

**O USO DE REDES NEURAIS ARTIFICIAIS NO
DIAGNÓSTICO PREDITIVO DOS TIPOS MAIS
FREQUENTES DE CEFALÉIA**

KARINA BORGES MENDES

**O USO DE REDES NEURAIS ARTIFICIAIS NO DIAGNÓSTICO PREDITIVO
DOS TIPOS MAIS FREQUENTES DE CEFALÉIA**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciências, do Programa de Pós-Graduação em Métodos Numéricos em Engenharia, na Área de Concentração em Programação Matemática, dos setores de Ciências Exatas e de Tecnologia da Universidade Federal do Paraná.

Orientadora: Prof^a Dr^a Maria Teresinha Arns
Steiner

CURITIBA
2008

À minha mãe Anelise

AGRADECIMENTOS

À minha mãe, por seu amor e apoio incondicionais.

À minha filha, Maria Antonia, razão de minha vida.

Aos meus avós, Oscarina e Nilson, exemplos de dignidade.

Ao meu padrasto, Fiuza, pela valiosa contribuição na coleta de dados para a realização dos experimentos e pelo conhecimento, incentivo e sugestões para o refinamento deste trabalho.

À minha orientadora, Profa. Maria Teresinha Arns Steiner, especialmente por sua ética e competência profissional.

Ao Prof. Jair Mendes Marques, pelas excepcionais contribuições.

Ao Prof. Alsemo Chaves Neto, pelo apoio na realização deste trabalho.

Ao Horácio, pela colaboração na elaboração dos algoritmos.

SUMÁRIO

LISTA DE FIGURAS.....	vi
LISTA DE QUADROS.....	vii
LISTA DE TABELAS.....	ix
RESUMO.....	x
ABSTRACT.....	xi
 CAPÍTULO I	
1 INTRODUÇÃO.....	1
1.1 OBJETIVOS DO TRABALHO.....	2
1.2 REVISÃO DA LITERATURA – TRABALHOS CORRELATOS.....	3
1.3 ESTRUTURA DO TRABALHO.....	7
 CAPÍTULO II	
2 DESCRIÇÃO DO PROBLEMA.....	9
2.1 INTRODUÇÃO.....	9
2.2 OBTENÇÃO DOS DADOS.....	12
2.3 TRATAMENTO DOS DADOS.....	15
 CAPÍTULO III	
3 REFERENCIAL TEÓRICO.....	28
3.1 REDES NEURAIS ARTIFICIAIS.....	28
3.1.1 Histórico.....	29
3.1.2 Características Básicas de uma Rede Neural.....	30
3.1.3 Função de Ativação de uma Rede Neural.....	32
3.1.4 Treinamento de uma Rede Neural.....	33
3.1.5 Fluxo de Dados em uma Rede Neural.....	35
3.1.6 Modelos de Rede Neural.....	36
3.1.6.1 <i>Perceptron</i>	36
3.1.6.2 Redes Lineares.....	37
3.1.6.3 Redes de Múltiplas Camadas.....	38

CAPÍTULO IV

4 IMPLEMENTAÇÃO DA REDE NEURAL ARTIFICIAL E ANÁLISE DOS

RESULTADOS.....	45
4.1 IMPLEMENTAÇÃO DA REDE NEURAL ARTIFICIAL.....	45
4.1.1 Teste I.....	48
4.1.2 Teste II.....	50
4.1.3 Teste III.....	52
4.2 ANÁLISE DOS RESULTADOS.....	54
4.2.1 Teste I.....	56
4.2.2 Teste II.....	57
4.2.3 Teste III.....	59

CAPÍTULO V

5 CONSIDERAÇÕES FINAIS.....

5.1 CONCLUSÕES.....	61
5.2 SUGESTÕES PARA TRABALHOS FUTUROS.....	63

REFERÊNCIAS.....

ANEXOS.....

LISTA DE FIGURAS

FIGURA 3.1	REPRESENTAÇÃO BÁSICA DO NEURÔNIO BIOLÓGICO..	31
FIGURA 3.2	EXEMPLO DE NEURÔNIO ARTIFICIAL.....	32
FIGURA 3.3	FUNÇÕES DE ATIVAÇÃO.....	33
FIGURA 3.4	ALGORITMO GENÉRICO DE UMA RNA SUPERVISIONADA.....	35
FIGURA 3.5	SEPARABILIDADE LINEAR.....	36
FIGURA 3.6	ILUSTRAÇÃO DE UMA REDE DE MÚLTIPLAS CAMADAS..	39

LISTA DE QUADROS

QUADRO 2.1	CODIFICAÇÃO DA VARIÁVEL “SEXO”	16
QUADRO 2.2	CODIFICAÇÃO DA VARIÁVEL “IDADE”	16
QUADRO 2.3	CODIFICAÇÃO DA VARIÁVEL “INÍCIO DA DOR”	16
QUADRO 2.4	CODIFICAÇÃO DA VARIÁVEL “LOCALIZAÇÃO DA DOR”	17
QUADRO 2.5	CODIFICAÇÃO DA VARIÁVEL “INTENSIDADE DA DOR”	18
QUADRO 2.6	CODIFICAÇÃO DA VARIÁVEL “TIPO DE DOR”	18
QUADRO 2.7	CODIFICAÇÃO DA VARIÁVEL “SURGIMENTO DA DOR”	19
QUADRO 2.8	CODIFICAÇÃO DA VARIÁVEL “EVOLUÇÃO DA DOR”	19
QUADRO 2.9	CODIFICAÇÃO DA VARIÁVEL “FREQÜÊNCIA DA DOR”	19
QUADRO 2.10	CODIFICAÇÃO DA VARIÁVEL “DURAÇÃO DO EPISÓDIO”	20
QUADRO 2.11	CODIFICAÇÃO DA VARIÁVEL “SINTOMAS ASSOCIADOS”	20
QUADRO 2.12	CODIFICAÇÃO DA VARIÁVEL “O QUE CAUSA OU PIORA A DOR”	22
QUADRO 2.13	CODIFICAÇÃO DA VARIÁVEL “O QUE MELHORA A DOR”	25
QUADRO 2.14	CODIFICAÇÃO DA VARIÁVEL “USO DE ANALGÉSICOS”	25
QUADRO 2.15	CODIFICAÇÃO DOS DADOS DE SAÍDA DA REDE.....	26
QUADRO 4.1	MATRIZ DE CONFUSÃO.....	55
QUADRO 4.2	MATRIZ DE CONFUSÃO – TESTE I.....	56
QUADRO 4.3	MATRIZ DE CONFUSÃO – RN I – TESTE II.....	57
QUADRO 4.4	MATRIZ DE CONFUSÃO – RN II – TESTE II.....	58

QUADRO 4.5	MATRIZ DE CONFUSÃO – RN III – TESTE II.....	58
QUADRO 4.6	MATRIZ DE CONFUSÃO – RN IV – TESTE II.....	58
QUADRO 4.7	MATRIZ DE CONFUSÃO – RN V – TESTE II.....	58
QUADRO 4.8	MATRIZ DE CONFUSÃO – TESTE III.....	59

LISTA DE TABELAS

TABELA 4.1	RESULTADOS ENCONTRADOS NO TESTE I.....	49
TABELA 4.2	RESULTADOS ENCONTRADOS NO TESTE II.....	51
TABELA 4.3	MELHORES RESULTADOS TESTE II.....	52
TABELA 4.4	RESULTADOS ENCONTRADOS NO TESTE III.....	53
TABELA 4.5	MELHORES RESULTADOS DAS REDES NEURAIS.....	54
TABELA 4.6	TAXA APARENTE DE ERRO.....	59

RESUMO

A dor de cabeça (cefaléia) é uma doença muito mais comum do que se imagina. Estudos mostram que chega a afetar cerca de 92% da população. Trata-se de uma doença de alto custo pessoal, social e econômico. Este trabalho tem como objetivo analisar a relevância dos dados coletados através de questionários respondidos por 2177 pacientes com diagnóstico de cefaléia e, através desta análise, verificar se o tratamento desses dados usando Redes Neurais Artificiais (RNAs) como ferramenta de reconhecimento de padrões pode auxiliar nos diagnósticos de novos pacientes. O sistema desenvolvido é baseado em RNAs do tipo *Perceptron* multicamadas e utilizou-se o software MATLAB 7.0 e o componente *Neural Network Toolbox* para a sua implementação e seu treinamento. No presente trabalho, são levantados elementos para justificar a utilização de RNAs como ferramenta de apoio ao diagnóstico, objetivando auxiliar o médico no seu dia-a-dia, e também como uma ferramenta educacional de auxílio ao treinamento e qualificação de profissionais da área médica. Os resultados obtidos foram bastante satisfatórios, mostrando que as RNAs podem ser eficientes na resolução deste problema específico.

ABSTRACT

Headache is a very common disease, even more common than we can think about. Studies show that it affects about 92% of the population. It is a high cost disease, related to social and economical status. The aim of this paper is to analyze the relevant results through questionnaires answered by 2177 patients with headache diagnosis, and check if the treatment, using Artificial Neural Networks (ANN) as a recognizing tool of patterns can help on new patients diagnosis. The developed system is based on Perceptron ANN multilayers and it was used MATLAB 7.0 software, and Neural Network Toolbox for its implementation and training. At this present paper, elements are studied to justify the use of ANN as a diagnosis tool support, as a way to help the doctor's routines, and as well as an educational tool on training qualifying medical professionals. The obtained results were satisfying, showing that ANN can be efficient on the resolution of this specific issue.

CAPÍTULO I

1 INTRODUÇÃO

Redes Neurais Artificiais (RNAs) são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

As RNAs são consideradas uma metodologia geral e prática para resolver problemas de Inteligência Artificial.

Reconhecimento de Padrões é uma das principais aplicações de RNAs, incluindo problemas em áreas como comércio, indústria, serviços, ensino, ciência, entre outras.

Na área médica, as RNAs têm sido usadas como suporte tanto para diagnóstico quanto para tratamento das mais variadas doenças.

As dores de cabeça (cefaléias) constituem um dos diagnósticos mais freqüentes na medicina. A cefaléia tem uma prevalência de 92% e corresponde a uma das maiores causas de consulta médica em qualquer serviço de saúde. Constitui, ainda, uma das causas mais freqüentes de absenteísmo e improdutividade no trabalho e escola.

O diagnóstico das cefaléias se faz através de consultas médicas, utilizando serviços de generalistas (clínicos, pediatras, ginecologistas, etc.) ou de especialistas (neurologistas, neurocirurgiões ou médicos dedicados exclusivamente a esta área de atuação, chamados cefaliatras).

A técnica utilizada para o diagnóstico consiste no interrogatório do paciente (anamnese), em seu exame clínico e em exames complementares. Atualmente tem importância crescente a utilização, na anamnese, de questionários estruturados ou semi-estruturados que possibilitem ao profissional maior abrangência no interrogatório e que facilitem o tratamento estatístico e a informatização desses dados.

A Sociedade Internacional de Cefaléia (*Internacional Headache Society - IHS*) descreve mais de 200 diagnósticos de dores de cabeça. Na prática médica cotidiana, é mais utilizada a Classificação Internacional das Doenças (CID-10) da Organização Mundial de Saúde (OMS). Essa classificação tem a possibilidade de ser usada de forma detalhada, como usualmente é feita por médicos especialistas. No presente trabalho optou-se pela classificação simplificada, usualmente mais utilizada por médicos generalistas.

Nos questionários semi-estruturados citados, para cada característica de um determinado sintoma são colocadas múltiplas possibilidades. Para se chegar a um diagnóstico utilizam-se diferentes combinações das características observadas pelo paciente em relação a cada sintoma.

Este trabalho busca analisar a relevância dos dados dos questionários coletados de 2177 pacientes com diagnóstico de cefaléia e, através desta análise, verificar se o tratamento desses dados pode auxiliar nos diagnósticos de novos pacientes.

1.1 OBJETIVOS DO TRABALHO

O objetivo deste trabalho é verificar se a aplicação de Redes Neurais Artificiais como técnica de reconhecimento de padrões pode auxiliar o diagnóstico médico dos tipos mais freqüentes de cefaléia.

A característica mais importante de uma Rede Neural Artificial é a sua capacidade de aprender e melhorar seu desempenho. Uma RNA baseia-se em dados reais para extrair um modelo geral, na tentativa de construir padrões detectados nesses dados (TATIBANA, 2006).

A implementação correta da Rede Neural Artificial para a questão das cefaléias pode apresentar as seguintes vantagens:

- Método de auxílio diagnóstico para orientação de médicos generalistas e profissionais que atuem em atenção primária à saúde;
- Método para agilizar e complementar o trabalho do médico especialista;
- Menor subjetividade no diagnóstico;
- Direcionamento mais eficaz do tratamento.

Como as RNAs aprendem através de exemplos, os detalhes de como reconhecer a doença, conhecendo-as mais profundamente, não são necessários ao usuário da RNA. Uma série de exemplos representativos de todas as variações da doença são necessários para a utilização de uma RNA (STERGIOUS, 2006).

1.2 REVISÃO DA LITERATURA – TRABALHOS CORRELATOS

Nesta seção é feita uma revisão da bibliografia, relatando algumas aplicações das RNAs nas mais diversas áreas.

PAVANELLI (2007) usou com sucesso as Redes Neurais Artificiais como ferramenta no reconhecimento de padrões para fornecer uma previsão do tempo de duração de um processo trabalhista aos usuários da Justiça do Trabalho. O autor utilizou 10 atributos de cada processo como entradas da RN. Os dados de entrada foram codificados de forma a tornarem-se binários. A função de ativação usada foi a função logística sigmoidal. Os 100 processos

que compunham a amostra foram divididos da seguinte forma: os dados de 75 processos foram usados para treinar a rede e os dados dos 25 processos restantes para testá-la.

BAPTISTELLA (2005) comparou as técnicas de Redes Neurais Artificiais e Regressão Linear Múltipla, na estimação dos valores venais de imóveis urbanos da cidade de Guarapuava / PR. O modelo inicial foi composto pelas 13 variáveis do Cadastro Imobiliário, fornecido pelo setor de Planejamento da Prefeitura Municipal de Guarapuava. As RNAs desenvolvidas foram do tipo *feed-forward*, utilizando o algoritmo de treinamento “Levenberg Marquardt” com uma camada oculta. Os resultados foram comparados com os resultados obtidos através da Regressão Linear Múltipla, para a predição dos valores venais. As duas técnicas se mostraram eficientes, sendo que, para o problema estudado, o desempenho das Redes Neurais Artificiais foi superior ao resultado apresentado pela outra técnica.

BALDIN (2002) aplicou a teoria e os algoritmos de RNAs como ferramenta auxiliar na Orientação Profissional de jovens. Utilizou os dados de 122 questionários respondidos por profissionais satisfeitos em sua área de atuação: área biológica, área humana ou área tecnológica. A RNA foi programada em Visual Basic. A camada de entrada foi composta por 10 neurônios, a camada escondida foi composta por neurônios variando de 1 a 10 e a camada de saída foi composta por um neurônio dicotômico numa primeira topologia, um neurônio tricotômico numa segunda topologia e três neurônios na terceira topologia. O melhor desempenho foi obtido usando-se a segunda topologia, um neurônio tricotômico na camada de saída.

LEMOS (2003) empregou Redes Neurais e Árvores de Decisão para auxiliar na tomada de decisão sobre conceder ou não crédito bancário a um cliente. Obteve resultados satisfatórios, sendo que usando RNs o percentual de erro foi menor que o obtido usando Árvores de Decisão. O treinamento da RNA foi feito considerando uma rede de múltiplas camadas usando o algoritmo *back-propagation*, variando os seguintes parâmetros: número de ciclos, conjunto de pesos iniciais e número de neurônios na camada escondida. Em

todos os testes a taxa de aprendizagem inicial foi igual a 0,01. A camada de entrada foi composta por 29 neurônios, referentes às informações obtidas de 339 empresas, das quais 266 eram adimplentes e as outras 73 inadimplentes.

ADAMOWICZ (2000) comparou duas técnicas de Reconhecimento de Padrões para discriminar empresas solventes de insolventes. Utilizou os seguintes métodos: um estatístico, Análise Discriminante Linear de Fisher e um matemático, RNAs. O conjunto de entrada é composto por nove índices econômico-financeiros de 136 empresas, das quais 118 foram pré-classificadas como solventes e 18 como insolventes. O desempenho das RNAs foi melhor que a da Análise Discriminante de Fisher, chegando a obter um percentual nulo de erro tanto para o conjunto de treinamento quanto para o conjunto de erro.

VIEIRA *et al* (2005) descrevem o processo de implementação de uma RNA como mecanismo de análise de similaridade de textos com retro-alimentação para um sistema de recuperação de informação na *web*. Como as RNAs têm suas entradas numéricas, criou-se uma representação para os termos na rede. Assim, cada entrada foi associada a um determinado termo, o que tornou necessária a elaboração de um dicionário global, responsável pela criação e ordenação de cada vetor de termos dos documentos em todas as categorias. Na camada de entrada o número de neurônios é igual ao número de termos do vetor de documentos. A camada intermediária tem 10 neurônios e a de saída apenas um, o qual recebe um valor no intervalo [0,1]. Porém, para representar o percentual de relevância do documento, multiplica-se o valor de saída por 100.

LIMA (2006) utilizou as RNAs como sistemas de aprendizado, simulação e otimização de processos de deposição de filmes finos poliméricos. Foram construídas redes neurais artificiais do tipo multilayer perceptron (MLP) utilizando o *software* MATLAB e o componente *Neural Networks Toolbox*. A reprodutibilidade e o número de neurônios contidos na camada intermediária foram avaliados. Os resultados mostraram que as RNAs fornecem boas respostas , mostrando que é possível a utilização dessa ferramenta para

auxiliar a engenharia de processos, as técnicas e análises de deposições de filmes finos poliméricos.

STEINER (1995) mostrou a utilização de RNAs no diagnóstico médico detectando, com uma certa probabilidade, se determinados pacientes tinham câncer ou cálculo no duto biliar. Para a construção da RNA, além da autora ter programado na linguagem Pascal o treinamento de uma RNA de Múltiplas Camadas, foi também utilizado o pacote computacional MATLAB – *Neural Network Toolbox*. Foram utilizados dados de 118 pacientes, sendo 35 portadores de câncer e 83 de cálculo. Ao programa foram apresentadas 14 variáveis sugeridas pelo médico especialista na área como entradas da RNA.

PRECHELT (1994) utilizou as RNAs para o diagnóstico médico de câncer de mama buscando classificar o tumor em maligno ou benigno baseado nas descrições celulares recolhidas pelo exame microscópico. Os atributos de entrada usados foram a espessura do grupo, a uniformidade do tamanho e da forma das células, a quantidade de adesão marginal, entre outros, num total de nove. Usou dois atributos de saída (tumor maligno ou tumor benigno). Das 699 pacientes, 65,5% eram portadoras de tumor benigno.

MASSAD *et al* (2004) fazem uso de RNAs no auxílio ao diagnóstico médico da doença de Alzheimer. Para isso, utilizam dados coletados pelo sistema computadorizado de uma avaliação neuropsicológica considerando cinco variáveis. As entradas são as porcentagens de acerto nos testes de reconhecimento de faces, desenhos, palavras, números e tempo de reação no teste dos desenhos. Da população estudada, 45 pacientes eram portadores da doença de Alzheimer e outros 45 pacientes eram normais. Dentre estes, os dados dos 30 indivíduos normais e 30 com a doença foram usados para treinar a RN e os dados de 15 pacientes normais e 15 com a doença foram usados para testar a validade do modelo utilizado.

MARTINS *et al* (2004) utilizaram RNAs com aprendizado não supervisionado (Rede de KOHONEN) para modelar o diagnóstico diferencial da doença meningocócica. Dos 100 casos utilizados, 65 foram para o treinamento

das redes e 35 para testa-la. A camada de entrada foi formada pelos dados clínicos dos pacientes num total de 34 neurônios, a camada intermediária pelos resultados laboratoriais e a camada de saída foi composta pelos oito diagnósticos possíveis. As Redes de Kohonen apresentaram um percentual de 83% de acerto, considerado bastante alto comparado aos resultados obtidos pelos clínicos.

SANTOS *et al* (2005) utilizaram as RNAs como instrumento de predição de tuberculose pulmonar em diversas Unidades de Saúde da Cidade do Rio de Janeiro. Dos 136 pacientes analisados, 43% apresentaram a Tuberculose Pulmonar Paucibacilar em atividade. Foram consideradas 12 variáveis de entrada tais como: idade, tosse, escarro, sudorese, entre outras. Diversas RNs foram treinadas considerando uma única saída ("1", se o paciente é portador da doença e "-1", se o paciente não é portador). O número de neurônios na camada intermediária foi escolhido experimentalmente, buscando um número pequeno, mas que proporcionasse uma boa generalização da rede. O melhor resultado (76% de acerto no diagnóstico) foi obtido numa RN com 12 neurônios na camada de entrada, cinco na camada intermediária e um na camada de saída.

1.3 ESTRUTURA DO TRABALHO

O presente trabalho está estruturado em cinco capítulos.

O primeiro capítulo trata da introdução ao trabalho. A descrição do problema é tratada no Capítulo II.

O Capítulo III descreve, em detalhes, os conceitos das Redes Neurais Artificiais.

O Capítulo IV descreve a implementação da Rede Neural e a análise dos resultados obtidos pela rede.

Finalmente, no Capítulo V encontram-se as conclusões obtidas pela análise dos resultados do Capítulo IV e as sugestões para futuros trabalhos.

CAPÍTULO II

2 DESCRIÇÃO DO PROBLEMA

2.1 INTRODUÇÃO

A dor de cabeça é uma causa maior de sofrimento humano e motivo definido de preocupação desde os primórdios da civilização. Há textos assírios e babilônios referentes ao problema, papiros egípcios com ilustrações e detalhada literatura médica proveniente da Grécia clássica, quando Hipócrates descreveu o primeiro caso de enxaqueca (MARANHÃO, 2002).

As cefaléias são muito freqüentes, com alguns estudos epidemiológicos estimando até 92% de prevalência. É, portanto, queixa que atinge a maioria das pessoas, sendo um dos maiores motivos de consulta médica (STOVNER *et al*, 2006). Isto gera impacto significativo na qualidade de vida do indivíduo, com interferência em suas relações no trabalho, família e lazer (DAHLÖF *et al*, 2006). Representa ainda ônus econômico e social considerável, tendo em vista não apenas os custos diretos no sistema de saúde, mas também os custos indiretos, vinculados ao absenteísmo e à perda de produtividade na escola ou trabalho (BERG *et al*, 2006).

Entre todos os diagnósticos de cefaléia definidos pela *IHS*, a cefaléia tensional e a enxaqueca são as mais freqüentes, sendo responsáveis pela maioria das consultas. Foram, por isto, levadas em consideração no presente trabalho, onde se dividiu a enxaqueca nos seus tipos principais: com e sem aura. Incluiu-se ainda a cefaléia por uso abusivo de medicamentos, por ser frequentemente uma evolução dos diagnósticos precedentes, consequência principalmente de tratamento inadequado.

A enxaqueca, segundo a IHS, é uma doença caracterizada por crises de forte dor de cabeça localizadas na metade da cabeça (hemicrânia), de característica latejante ou pulsátil, de duração de 4 a 72 horas, acompanhadas de náuseas, vômitos, intolerância à luz (fotofobia) e ao barulho (fonofobia). Outras doenças mais graves devem ser afastadas pelo exame físico ou exames complementares. A anamnese que contemplar a maior parte destes dados torna as queixas do paciente compatíveis com o diagnóstico de enxaqueca. Se as crises forem precedidas por sintomas visuais ou queixas em outros órgãos dos sentidos, caracteriza-se a enxaqueca com aura. Caso contrário, o diagnóstico é enxaqueca sem aura. A enxaqueca atinge 18% das mulheres e 6% dos homens.

A cefaléia tensional é caracterizada por episódios de dor de cabeça bilateral, em peso ou pressão, de intensidade moderada, sem a presença de vômitos ou outros sintomas comuns à enxaqueca. É a cefaléia mais comum. Alguns levantamentos indicam prevalência acima de 80%.

A cefaléia por uso abusivo de medicamentos (CEM) se diferencia essencialmente pelo fato de o paciente fazer uso de medicações analgésicas abusivamente, acima de 15 dias por mês. Este quadro atinge 3% das pessoas e é muito freqüente nos serviços especializados em cefaléia.

Para a orientação adequada relativa aos seus sintomas, o paciente deve procurar o sistema de saúde, para submeter-se a consulta médica, quando o profissional, especialista ou não, procurará instituir primeiramente um diagnóstico adequado, após o qual o tratamento será instituído.

A técnica utilizada para o diagnóstico consiste inicialmente no interrogatório para investigar a história da moléstia do paciente (anamnese). É então importante qualificar e quantificar o sintoma, levantando-se a duração da queixa e idade do paciente no início da mesma, a localização e duração dos ataques, bem como a sua freqüência, a característica e intensidade da dor, os fatores que a precipitam e os que a aliviam (incluindo medicações) e ainda os fatores que acompanham o fenômeno doloroso. Atualmente tem importância

crescente a utilização de questionários estruturados ou semi-estruturados, que possibilitam ao profissional maior abrangência no interrogatório e facilitam o tratamento estatístico e a informatização desses dados (OLESEN *et al*, 2006). Após este interrogatório dirigido à cefaléia, são feitas perguntas sobre outras doenças atuais e passadas, sobre o estado psicológico, doenças familiares e hábitos. O exame físico completa a consulta médica, sendo então postulado o diagnóstico, solicitados exames e feitas orientações iniciais.

A *IHS* descreve com pormenores mais de 200 doenças capazes de causar dores de cabeça. O detalhamento desta classificação permitiu a unificação da linguagem dos especialistas da área, acarretando grande desenvolvimento na clínica e pesquisa das cefaléias. Na prática médica cotidiana, é mais utilizada a Classificação Internacional das Doenças (CID-10) da Organização Mundial de Saúde (OMS), cujas principais diretrizes são derivadas da classificação anterior (*IHS*). Essa classificação tem a possibilidade de ser usada de forma detalhada, com subdivisões de até 5 dígitos, como usualmente é feita por médicos especialistas. Na atividade clínica rotineira é freqüente a opção pela classificação simplificada, com três dígitos, usualmente mais utilizada por médicos generalistas.

Entretanto, a maioria dos pacientes com cefaléia não procura ou não tem acesso ao sistema de saúde, recorrendo à auto-medicação. Entre os que procuram o médico, a maior parte é atendida pelo sistema primário de saúde (clínicos gerais), não sendo examinados pelos níveis secundário ou terciário (especialistas). Como consequência destas distorções, estima-se que cerca da metade dos pacientes nunca receba um diagnóstico correto e que, por isto, não sejam tratados adequadamente (LIPTON *et al*, 2007).

Para melhorar esta situação é necessário aumentar a proporção de pacientes com acesso a diagnóstico e tratamento adequado. Como recomenda a OMS, é possível desenvolver um sistema de complexidade crescente, prevendo amplo acesso à porta de entrada e encaminhamento dos casos mais complexos. O primeiro passo é a entrevista do paciente com profissional de

saúde (não médico) treinado para aplicar questionários preditivos de diagnóstico provisório de cefaléia.

Com a validação de um questionário e o treinamento de agentes de saúde, o acesso a este atendimento pode ser muito amplo. A partir da discriminação resultante do diagnóstico provisório, os pacientes podem ser encaminhados ao nível de atendimento adequado. Os casos simples podem ser orientados pelo próprio agente de saúde (há diretrizes para isto). Os casos necessitando de orientação médica passam pelo médico de família ou clínico geral. Os casos complexos são encaminhados ao especialista ou mesmo à internação hospitalar.

Há diversos tipos de questionários informatizados desenvolvidos para funcionar como *screening* (BENSEÑOR *et al*, 1997 e RASMUNSEN, 1991). Os trabalhos até agora feitos utilizam as informações fornecidas pela Classificação Internacional das Cefaléias para o levantamento das questões essenciais ao diagnóstico. Algoritmos desenvolvidos indicam a doença provável. A maioria visa apontar os quadros mais freqüentes, entre os quais se destacam: enxaqueca nos seus tipos mais comuns (com e sem aura), cefaléia tipo tensional e cefaléia por uso excessivo de medicação (BAREA, 2002).

A idéia central do presente trabalho é utilizar Redes Neurais Artificiais para o tratamento dos dados obtidos de questionários de pacientes com cefaléia, na expectativa do desenvolvimento de um sistema alternativo aos já existentes, podendo ser mais flexível, com maiores possibilidades adaptativas e que possa espelhar melhor o raciocínio de um profissional da área.

2.2 OBTENÇÃO DOS DADOS

Os dados utilizados neste trabalho foram extraídos do banco de dados da Clínica Neurológica de Joinville/ SC, através de um questionário aplicado a 2177 pacientes com diagnóstico de cefaléia, de acordo com o ANEXO I deste trabalho. As perguntas incluídas no questionário formam um corpo capaz de

preencher os critérios necessários para o diagnóstico das principais causas de cefaléia, como orienta a *IHS*. Na rotina da Clínica Neurológica de Joinville, o questionário é utilizado apenas para a obtenção de informações. Os dados não tinham sido até hoje tratados por nenhum tipo de algoritmo visando diagnóstico ou qualquer outro propósito. Os diagnósticos destes pacientes foram feitos a partir da experiência clínica de um único médico, que milita na área de atuação das dores de cabeça: cefaliatria. Os dados retratam a evolução de pacientes de janeiro de 2002 a novembro de 2006. Não foram consideradas possíveis mudanças de diagnóstico dos pacientes após novembro de 2006. Os nomes dos pacientes foram omitidos por questão de sigilo.

Dos 2177 pacientes com cefaléia, 1252 pacientes tiveram o diagnóstico de cefaléia tensional, 307 pacientes de enxaqueca sem aura, 99 pacientes de enxaqueca com aura, 100 pacientes de cefaléia por uso excessivo de medicamento e 419 pacientes tiveram outros tipos de diagnóstico.

O questionário constitui-se de 14 informações sobre cada um dos pacientes atendidos durante o período anteriormente citado e encontram-se disponíveis no ANEXO I deste trabalho. É importante salientar que a identidade dos pacientes foi preservada, por uma questão de ética e sigilo profissional.

As 14 perguntas do questionário que constituem as variáveis de entrada do problema são apresentadas a seguir:

1. Sexo: feminino ou masculino;
2. Idade: até 12 anos, de 12 a 19 anos, de 20 a 39 anos ou acima dos 40 anos;
3. Início da dor: dias, semanas, meses e/ou anos;
4. Localização da dor: hemicrânia, holocraniana, frontal, occipital e/ou bem localizada;

5. Intensidade da dor: forte, muito forte, moderada e/ou fraca;
6. Tipo de dor: latejante, peso, pontada e/ou atípica;
7. Surgimento da dor: crise, contínua, episódica e/ou em salvas;
8. Evolução da dor: estável, mudando e/ou progressiva;
9. Frequência da dor: diária, mais de 15 episódios ao mês, 1 a 2 episódios por semana, 1 a 3 episódios ao mês e/ou alguns episódios ao ano;
10. Duração do Episódio: segundos, minutos, horas, dias e/ou semanas;
11. Sintomas associados a dor: náusea, vômito, fotofobia, fonofobia, transtornos visuais, tontura e/ou febre;
12. O que causa ou piora a dor: nervosismo, menstruação, alguns alimentos, bebidas, sono irregular, esforço físico, sexo e/ou movimentos do pescoço;
13. O que melhora a dor: deitar ou dormir, relaxar, distrair-se, caminhadas e/ou gravidez;
14. Uso de analgésicos: nunca, raramente, sempre que dói, quase todos os dias, diariamente e/ou vários ao dia.

Pode-se constatar que as entradas dizem respeito aos dados pessoais do paciente e às características de seu sintoma, a dor de cabeça. As múltiplas alternativas de resposta podem ser escolhidas isoladamente ou de maneira múltipla, já que a dor pode ter característica variável.

2.3 TRATAMENTO DOS DADOS

A escolha e adequação dos dados utilizados para treinar e testar uma RNA é de fundamental importância. É necessário que se disponha de dados em quantidade e qualidade suficientes. Caso a quantidade de dados seja pequena, a rede não conseguirá criar um modelo suficientemente representativo para se ter um desempenho satisfatório quando aplicado em situações reais após o seu desenvolvimento. Além disto, os dados devem englobar todos os aspectos do problema em questão, a fim de que o modelo criado seja genérico. Em geral, tais dados precisam ser convertidos para um formato padrão para utilização pelas RNAs.

Com o objetivo de treinar uma RNA que forneça um diagnóstico preditivo dos tipos mais freqüentes de cefaléia e que apresente um bom desempenho, as entradas citadas em (2.2) foram “tratadas”, conforme comentado a seguir.

As variáveis explicativas são do tipo quantitativas (idade) e qualitativas (sexo, início, localização, intensidade, tipo, surgimento, evolução, freqüência, duração, sintomas, o que causa e o que melhora a dor e uso de analgésicos) e representam as características (sintomas) do paciente.

A codificação 1 define as entradas através de uma escala numérica, atribuindo pesos às modalidades, com valores de 1 a n , onde n representa o número de opções para cada sintoma. A codificação 2 “trata” as entradas transformando-as em coordenadas binárias, atribuindo o valor “0” quando o paciente não possui a característica e o valor “1” caso contrário.

A variável “Sexo” apresentará na codificação 1 o valor “1” se o paciente for mulher e “2” se for homem. Na codificação 2, o vetor (1,0) será a representação para o paciente do sexo feminino e (0,1) para o sexo masculino. Assim, o atributo “Sexo” representa um vetor com duas coordenadas binárias. O quadro 2.1 mostra a codificação da variável “Sexo”.

QUADRO 2.1 CODIFICAÇÃO DA VARIÁVEL “SEXO”

Variável Sexo	Codificação 1	Codificação 2	
		Entrada 1	Entrada 2
Sexo feminino	1	1	0
Sexo masculino	2	0	1

O atributo “Idade” foi dividido em faixas etárias (intervalos). Na codificação 1, o paciente com idade até 12 anos recebe o valor “1”, de 12 a 19 anos, o valor “2”, de 20 a 39 anos, o valor “3” e acima de 40 anos, o valor “4”. As coordenadas binárias são mostradas na codificação 2, representando quatro coordenadas no vetor de entrada da rede neural. O quadro 2.2 mostra as codificações 1 e 2 para a variável de entrada “Idade”.

QUADRO 2.2 CODIFICAÇÃO DA VARIÁVEL “IDADE”

Variável Idade	Codificação 1	Codificação 2			
		Entrada 1	Entrada 2	Entrada 3	Entrada 4
Até 12 anos	1	1	0	0	0
12 a 19 anos	2	0	1	0	0
20 a 39 anos	3	0	0	1	0
Acima de 40 anos	4	0	0	0	1

Com relação ao “Início da dor”, o atributo foi codificado como mostra o quadro 2.3 a seguir, contando com quatro coordenadas no vetor de entrada de dados na Codificação 2. Para este atributo pode existir mais de uma resposta, logo o número de opções de resposta para os 2177 pacientes passou de quatro para seis, como mostra a Codificação 1.

QUADRO 2.3 CODIFICAÇÃO DA VARIÁVEL “INÍCIO DA DOR”

Variável Início da dor	Codificação 1	Codificação 2			
		Entrada 1	Entrada 2	Entrada 3	Entrada 4
Dias	1	1	0	0	0
Semanas	2	0	1	0	0
Meses	3	0	0	1	0
Anos	4	0	0	0	1
Semanas + Meses	5	0	1	1	0
Meses + Anos	6	0	0	1	1

A variável “Localização da dor” tem codificação conforme o quadro 2.4, mostrado a seguir. Esta variável possui cinco coordenadas no vetor de entrada de dados na Codificação 2. É importante salientar que a cefaléia nem sempre

tem localização unilateral, podendo manifestar-se bilateralmente. Sendo assim, o número de combinações encontradas nos 2177 pacientes em relação ao atributo “Localização da dor” foi igual a 21 (Codificação 1).

QUADRO 2.4 CODIFICAÇÃO DA VARIÁVEL “LOCALIZAÇÃO DA DOR”

Variável Localização da dor	Codificação 1	Codificação 2				
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5
Hemicrânia	1	1	0	0	0	0
Holocraniana	2	0	1	0	0	0
Frontal	3	0	0	1	0	0
Occipital	4	0	0	0	1	0
Bem localizada	5	0	0	0	0	1
Hemicrânia + Holocraniana	6	1	1	0	0	0
Hemicrânia + Holocraniana + Frontal	7	1	1	1	0	0
Holocraniana + Frontal	8	0	1	1	0	0
Hemicrânia + Frontal	9	1	0	1	0	0
Hemicrânia + Holocraniana + Frontal + Occipital	10	1	1	1	1	0
Holocraniana + Frontal + Occipital	11	0	1	1	1	0
Hemicrânia + Frontal + Occipital	12	1	0	1	1	0
Frontal + Occipital	13	0	0	1	1	0
Hemicrânia + Holocraniana + Occipital	14	1	1	0	1	0
Holocraniana + Occipital	15	0	1	0	1	0
Hemicrânia + Occipital	16	1	0	0	1	0
Hemicrânia + Occipital + Bem localizada	17	1	0	0	1	1
Occipital + Bem localizada	18	0	0	0	1	1
Holocraniana + Bem localizada	19	0	1	0	0	1
Hemicrânia + Bem localizada	20	1	0	0	0	1
Bem localizada	21	0	0	0	0	1

O atributo “Intensidade da Dor” foi codificado segundo o quadro 2.5, a seguir. A cefaléia pode apresentar diferentes intensidades de dor, por isso, o número de combinações foi igual a 10. Na codificação 1, a intensidade forte recebe o valor “1”, muito forte, o valor “2”, moderada, o valor “3” e fraca, o valor “4”. A seguir a combinação forte e muito forte apresenta o valor “5” e assim por diante, até a última combinação de intensidade, que recebe o valor “10”. A codificação 2 apresenta um vetor de quatro coordenadas binárias. O quadro 2.5, abaixo, mostra as codificações.

QUADRO 2.5 CODIFICAÇÃO DA VARIÁVEL "INTENSIDADE DA DOR"

Variável Intensidade da dor	Codificação 1	Codificação 2			
		Entrada 1	Entrada 2	Entrada 3	Entrada 4
Forte	1	1	0	0	0
Muito forte	2	0	1	0	0
Moderada	3	0	0	1	0
Fraca	4	0	0	0	1
Forte + Muito Forte	5	1	1	0	0
Forte + Muito Forte + Moderada	6	1	1	1	0
Muito Forte + Moderada	7	0	1	1	0
Forte + Moderada	8	1	0	1	0
Forte + Moderada + Fraca	9	1	0	1	1
Moderada + Fraca	10	0	0	1	1

Para a variável "Tipo de Dor" a codificação encontra-se no quadro 2.6, mostrado abaixo. Na codificação 1 o número de opções de resposta é 15, já que o paciente pode ter mais de um tipo de dor. Na codificação 2, o vetor de entradas da rede tem quatro coordenadas binárias.

QUADRO 2.6 CODIFICAÇÃO DA VARIÁVEL "TIPO DE DOR"

Variável Intensidade da dor	Codificação 1	Codificação 2			
		Entrada 1	Entrada 2	Entrada 3	Entrada 4
Latejante	1	1	0	0	0
Peso	2	0	1	0	0
Pontada	3	0	0	1	0
Atípica	4	0	0	0	1
Latejante + Peso	5	1	1	0	0
Latejante + Peso + Pontada	6	1	1	1	0
Peso + Pontada	7	0	1	1	0
Latejante + Pontada	8	1	0	1	0
Latejante + Peso + Pontada + Atípica	9	1	1	1	1
Peso + Pontada + Atípica	10	0	1	1	1
Latejante + Pontada + Atípica	11	1	0	1	1
Pontada + Atípica	12	0	0	1	1
Latejante + Peso + Atípica	13	1	1	0	1
Peso + Atípica	14	0	1	0	1
Latejante + Atípica	15	1	0	0	1

Com relação ao atributo "Surgimento da Dor", o quadro 2.7 mostra as codificações 1 e 2. Na codificação 1 o número de combinações foi 6 e na codificação 2 o vetor de entradas é formado por 4 coordenadas binárias.

QUADRO 2.7 CODIFICAÇÃO DA VARIÁVEL “SURGIMENTO DA DOR”

Variável Surgimento da dor	Codificação 1	Codificação 2			
		Entrada 1	Entrada 2	Entrada 3	Entrada 4
Crise	1	1	0	0	0
Contínua	2	0	1	0	0
Episódica	3	0	0	1	0
Salvas	4	0	0	0	1
Crise + Contínua	5	1	1	0	0
Crise + Episódica	6	1	0	1	0

A variável “Evolução da Dor” tem suas codificações mostradas no quadro 2.8 a seguir. O atributo pode mostrar 3 estágios: estável, mudando ou progressiva. Na codificação 1, o valores vão de 1 a 3, respectivamente. Na codificação 2, o vetor de entradas da RNA foi representado por três coordenadas binárias.

QUADRO 2.8 CODIFICAÇÃO DA VARIÁVEL “EVOLUÇÃO DA DOR”

Variável Evolução da dor	Codificação 1	Codificação 2		
		Entrada 1	Entrada 2	Entrada 3
Estável	1	1	0	0
Mudando	2	0	1	0
Progressiva	3	0	0	1

O atributo “Frequência da Dor” foi codificado como mostra o quadro 2.9, a seguir.

QUADRO 2.9 CODIFICAÇÃO DA VARIÁVEL “FREQUÊNCIA DA DOR”

Variável Frequência da dor	Codificação 1	Codificação 2				
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5
Diária	1	1	0	0	0	0
Quase Diária	2	0	1	0	0	0
1 a 2 por semana	3	0	0	1	0	0
1 a 3 ao mês	4	0	0	0	1	0
Algumas ao ano	5	0	0	0	0	1
Diária + Quase Diária	6	1	1	0	0	0
Quase Diária + 1 a 2 por semana	7	0	1	1	0	0
Diária + 1 a 2 por semana	8	1	0	1	0	0
1 a 3 ao mês + Algumas ao ano	9	0	0	0	1	1

Para o atributo “Duração do Episódio” a codificação está no quadro 2.10.

QUADRO 2.10 CODIFICAÇÃO DA VARIÁVEL “DURAÇÃO DO EPISÓDIO”

Variável Duração do Episódio	Codificação 1	Codificação 2				
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5
Segundos	1	1	0	0	0	0
Minutos	2	0	1	0	0	0
Horas	3	0	0	1	0	0
Dias	4	0	0	0	1	0
Semanas	5	0	0	0	0	1
Segundos + Minutos	6	1	1	0	0	0
Minutos + Horas	7	0	1	1	0	0
Segundos + Horas	8	1	0	1	0	0
Horas + Dias	9	0	0	1	1	0
Horas + Dias + Semanas	10	0	0	1	1	1
Dias + Semanas	11	0	0	0	1	1

O quadro 2.11 mostra as codificações da variável “Sintomas Associados”.

QUADRO 2.11 CODIFICAÇÃO DA VARIÁVEL “SINTOMAS ASSOCIADOS”

Variável Sintomas associados	Codificação 1	Codificação 2						
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5	Entrada 6	Entrada 7
Náusea	1	1	0	0	0	0	0	0
Vômito	2	0	1	0	1	0	0	0
Fotofobia	3	0	0	1	0	0	0	0
Fonofobia	4	0	0	0	1	0	0	0
Transtornos visuais	5	0	0	0	0	1	0	0
Tontura	6	0	0	0	0	0	1	0
Febre	7	0	0	0	0	0	0	1
Náusea + Vômito	8	1	1	0	0	0	0	0
Náusea + Vômito + Fotofobia	9	1	1	1	0	0	0	0
Vômito + Fotofobia	10	0	1	1	0	0	0	0
Náusea + Fotofobia	11	1	0	1	0	0	0	0
Náusea + Vômito + Fotofobia + Fonofobia	12	1	1	1	1	0	0	0
Vômito + Fotofobia + Fonofobia	13	0	1	1	1	0	0	0
Náusea + Fotofobia + Fonofobia	14	1	0	1	1	0	0	0
Fotofobia + Fonofobia	15	0	0	1	1	0	0	0
Náusea + Vômito + Fonofobia	16	1	1	0	1	0	0	0
Vômito + Fonofobia	17	0	1	0	1	0	0	0
Náusea + Fonofobia	18	1	0	0	1	0	0	0
Náusea + Vômito + Fotofobia + Fonofobia + Transtornos Visuais	19	1	1	1	1	1	0	0
Vômito + Fotofobia + Fonofobia + Transtornos Visuais	20	0	1	1	1	1	0	0
Náusea + Fotofobia + Fonofobia + Transtornos Visuais	21	1	0	1	1	1	0	0
Fotofobia + Fonofobia + Transtornos Visuais	22	0	0	1	1	1	0	0
Náusea + Vômito + Fonofobia + Transtornos Visuais	23	1	1	0	1	1	0	0
Náusea + Fonofobia + Transtornos	24	1	0	0	1	1	0	0

Visuais								
Fonofobia + Transtornos Visuais	25	0	0	0	1	1	0	0
Náusea + Vômito + Fotofobia + Transtornos Visuais	26	1	1	1	0	1	0	0
Vômito + Fotofobia + Transtornos Visuais	27	0	1	1	0	1	0	0
Fotofobia + Transtornos Visuais	28	0	0	1	0	1	0	0
Vômito + Transtornos Visuais	29	0	1	0	0	1	0	0
Náusea + Transtornos Visuais	30	1	0	0	0	1	0	0
Náusea + Vômito + Fotofobia + Fonofobia + Transtornos Visuais + Tontura	31	1	1	1	1	1	1	0
Náusea + Fonofobia + Transtornos Visuais + Tontura	32	1	0	0	1	1	1	0
Fotofobia + Transtornos Visuais + Tontura	33	0	0	1	0	1	1	0
Náusea + Vômito + Fotofobia + Transtornos Visuais + Tontura	34	1	1	1	0	1	1	0
Fotofobia + Transtornos Visuais + Tontura	35	0	0	1	0	1	1	0
Náusea + Transtornos Visuais + Tontura	36	1	0	0	0	1	1	0
Transtornos Visuais + Tontura	37	0	0	0	0	1	1	0
Náusea + Vômito + Fotofobia + Fonofobia + Tontura	38	1	1	1	1	0	1	0
Náusea + Fotofobia + Fonofobia + Tontura	39	1	0	1	1	0	1	0
Fotofobia + Fonofobia + Tontura	40	0	0	1	1	0	1	0
Náusea + Vômito + Fonofobia + Tontura	41	1	1	0	1	0	1	0
Náusea + Fonofobia + Tontura	42	1	0	0	1	0	1	0
Fonofobia + Tontura	43	0	0	0	1	0	1	0
Náusea + Vômito + Fotofobia + Tontura	44	1	1	1	0	0	1	0
Náusea + Fotofobia + Tontura	45	1	0	1	0	0	1	0
Fotofobia + Tontura	46	0	0	1	0	0	1	0
Náusea + Vômito + Tontura	47	1	1	0	0	0	1	0
Vômito + Tontura	48	0	1	0	0	0	1	0
Náusea + Tontura	49	1	0	0	0	0	1	0
Náusea + Vômito + Tontura + Febre	50	1	1	0	0	0	1	1
Náusea + Tontura + Febre	51	1	0	0	0	0	1	1
Tontura + Febre	52	0	0	0	0	0	1	1
Náusea + Vômito + Febre	53	1	1	0	0	0	0	1
Vômito + Febre	54	0	1	0	0	0	0	1
Náusea + Febre	55	1	0	0	0	0	0	1

A variável “O que causa ou piora a dor” é codificada segundo o quadro 2.12, mostrado a seguir.

QUADRO 2.12 CODIFICAÇÃO DA VARIÁVEL “O QUE CAUSA OU PIORA A DOR”

Variável O que causa ou piora a dor	Codificação 1	Codificação 2							
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5	Entrada 6	Entrada 7	Entrada 8
Nervosismo	1	1	0	0	0	0	0	0	0
Menstruação	2	0	1	0	0	1	0	0	0
Alimentos	3	0	0	1	0	0	0	0	0
Bebidas	4	0	0	0	1	0	0	0	0
Sono	5	0	0	0	0	1	0	0	0
Esforço Físico	6	0	0	0	0	0	1	0	0
Sexo	7	0	0	0	0	0	0	1	0
Movimentos do Pescoço	8	0	0	0	0	0	0	0	1
Nervosismo + Menstruação	9	1	1	0	0	0	0	0	0
Nervosismo + Menstruação + Alimentos +	10	1	1	1	0	0	0	0	0
Menstruação + Alimentos	11	0	1	1	0	0	0	0	0
Nervosismo + Alimentos	12	1	0	1	0	0	0	0	0
Menstruação + Alimentos + Bebidas	13	0	1	1	1	0	0	0	0
Nervosismo + Alimentos + Bebidas	14	1	0	1	1	0	0	0	0
Alimentos + Bebidas	15	0	0	1	1	0	0	0	0
Nervosismo + Menstruação + Bebidas	16	1	0	1	1	0	0	0	0
Menstruação + Bebidas	17	0	1	0	1	0	0	0	0
Nervosismo + Bebidas	18	1	0	0	1	0	0	0	0
Menstruação + Alimentos + Bebidas + Sono	19	0	1	1	1	1	0	0	0
Nervosismo + Alimentos + Bebidas + Sono	20	1	0	1	1	1	0	0	0
Nervosismo + Bebidas + Sono	21	1	0	0	1	1	0	0	0
Bebidas + Sono	22	0	0	0	1	1	0	0	0
Nervosismo + Alimentos + Sono	23	1	0	1	0	1	0	0	0
Alimentos + Sono	24	0	0	1	0	1	0	0	0
Nervosismo + Menstruação + Sono	25	1	1	0	0	1	0	0	0
Menstruação + Sono	26	0	1	0	0	1	0	0	0
Nervosismo + Sono	27	1	0	0	0	1	0	0	0
Menstruação + Alimentos + Bebidas + Sono + Esforço	28	0	1	1	1	1	1	0	0
Bebidas + Sono + Esforço	29	0	0	0	1	1	1	0	0
Nervosismo + Alimentos + Sono + Esforço	30	1	0	1	0	1	1	0	0
Sono + Esforço	31	0	0	0	0	1	1	0	0
Menstruação +	32	0	1	1	1	0	1	0	0

Alimentos + Bebidas + Esforço									
Nervosismo + Bebidas + Esforço	33	1	0	0	1	0	1	0	0
Nervosismo + Menstruação + + Esforço	34	1	1	0	0	0	1	0	0
Menstruação + Esforço	35	0	1	0	0	0	1	0	0
Nervosismo + Esforço	36	1	0	0	0	0	1	0	0
Esforço + Sexo	37	0	0	0	0	0	1	1	0
Menstruação + Sexo	38	0	1	0	0	0	0	1	0
Nervosismo + Sexo	39	1	0	0	0	0	0	1	0
Esforço + Pescoço	40	0	0	0	0	0	1	0	1
Nervosismo + Menstruação + Sono + Pescoço	41	1	1	0	0	1	0	0	1
Nervosismo + Sono + Pescoço	42	1	0	0	0	1	0	0	1
Alimentos + Pescoço	43	0	0	1	0	0	0	0	1
Nervosismo + Pescoço	44	1	0	0	0	0	0	0	1

O quadro 2.13 a seguir mostra a codificação da variável “O que melhora a dor”.

QUADRO 2.13 CODIFICAÇÃO DA VARIÁVEL “O QUE MELHORA A DOR”

Variável O que melhora a dor	Codificação 1	Codificação 2				
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5
Deitar ou Dormir	1	1	0	0	0	0
Relaxar	2	0	1	0	0	0
Distrair-se	3	0	0	1	0	0
Caminhadas	4	0	0	0	1	0
Gravidez	5	0	0	0	0	1
Deitar + Relaxar	6	1	1	0	0	0
Distrair-se + Caminhadas	7	0	0	1	1	0
Relaxar + Caminhadas	8	0	1	0	1	0
Deitar + Gravidez	9	1	0	0	0	1

Finalmente, o quadro 2.14 de codificação do atributo “Uso de Analgésicos” é mostrado a seguir.

QUADRO 2.14 CODIFICAÇÃO DA VARIÁVEL “USO DE ANALGÉSICOS”

Variável Uso de Analgésicos	Codificação 1	Codificação 2					
		Entrada 1	Entrada 2	Entrada 3	Entrada 4	Entrada 5	Entrada 6
Nunca	1	1	0	0	0	0	0
Raramente	2	0	1	0	0	0	0
Sempre que dói	3	0	0	1	0	0	0
Quase diariamente	4	0	0	0	1	0	0
Diariamente	5	0	0	0	0	1	0
Vários ao dia	6	0	1	1	0	0	1
Sempre que dói + Quase diariamente	7	0	0	1	1	0	0
Sempre que dói + Diariamente	8	0	0	1	0	1	0
Diariamente + Vários ao dia	9	0	0	0	0	1	1
Quase diariamente + Vários ao dia	10	0	0	0	1	0	1
Sempre que dói + Vários ao dia	11	0	0	1	0	0	1

Na codificação 1, a matriz de entrada de dados é de ordem 2177 x 14 (ANEXO III).

Na codificação 2, tem-se um vetor de entradas com 66 coordenadas binárias, que correspondem aos 14 atributos de cada paciente. A matriz de entrada de dados para a codificação 2, disponível no ANEXO IV, é de ordem 2177 x 66, já que foram considerados os dados referentes a 2177 pacientes.

A variável resposta (saída da RNA) é o diagnóstico do paciente. Em uma primeira etapa, foram consideradas cinco saídas para a RNA: Cefaléia Tensional, Enxaqueca sem Aura, Enxaqueca com Aura, Cefaléia por Uso Excessivo de Medicamento e Outros Tipos de Cefaléia, cada uma delas correspondendo a um dos diagnósticos de cefaléia. O quadro 2.15 a seguir traz a codificação dos dados de saída.

QUADRO 2.15 CODIFICAÇÃO DOS DADOS DE SAÍDA DA REDE

Saídas da Rede	Codificação				
	Saída 1	Saída 2	Saída 3	Saída 4	Saída 5
Cefaléia Tensional	1	0	0	0	0
Enxaqueca sem Aura	0	1	0	0	0
Enxaqueca com Aura	0	0	1	0	0
Cefaléia por Uso Excessivo de Medicamento	0	0	0	1	0
Outros tipos de cefaléia	0	0	0	0	1

Numa segunda etapa foram consideradas cinco RNAs, cada uma delas para fazer o treinamento de cada um dos diagnósticos citados no quadro 2.15 anterior, com apenas uma saída do tipo “0” ou “1”.

A primeira rede tem como saída o diagnóstico “Cefaléia Tensional”. Assim, o paciente com este diagnóstico apresenta o valor atribuído “1”; caso contrário, o paciente apresenta o valor atribuído “0”.

O mesmo critério foi adotado para as quatro demais redes treinadas. São elas:

- RNA I: Cefaléia Tensional;
- RNA II: Enxaqueca sem Aura;
- RNA III: Cefaléia por Uso Excessivo de Medicamentos;
- RNA IV: Enxaqueca com Aura;
- RNA V: Outros Tipos de Cefaléia.

A codificação dos valores de saída (quadro 2.15) podem ser vistas no ANEXO V.

Assim, as variáveis originais entrada e saída foram codificadas adequadamente para que a implementação das RNAs possam ser realizadas com sucesso.

CAPÍTULO III

3 REFERENCIAL TEÓRICO

3.1 REDES NEURAIS ARTIFICIAIS

Uma Rede Neural Artificial é uma técnica com muitos processadores simples (cada um deles tendo, possivelmente, uma pequena quantidade de memória local), conectados por meio de canais de comunicação (conexões), aos quais, usualmente, estão associados valores (pesos) numéricos (PALMA NETO, 2005).

Assim, uma RNA pode ser caracterizada:

- por seus processadores básicos (neurônios);
- pela função de ativação;
- pelo padrão de conexão existente entre os neurônios;
- pelo algoritmo de aprendizagem.

RNAs são um conceito da computação que visa trabalhar no processamento de dados de maneira semelhante ao cérebro humano. O cérebro é tido como um processador altamente complexo e que realiza processamentos de maneira paralela. Para isso, ele organiza sua estrutura, ou seja, os neurônios, de forma que eles realizem o processamento necessário. Isso é feito numa velocidade extremamente alta e não existe qualquer computador no mundo capaz de realizar o que o cérebro humano faz.

Nas RNAs, a idéia é realizar o processamento de informações tendo como princípio a organização de neurônios do cérebro. Como o cérebro humano é capaz de aprender e tomar decisões baseadas na aprendizagem, as redes neurais artificiais devem fazer o mesmo. Assim, uma RNA pode ser interpretada como um esquema de processamento capaz de armazenar

conhecimento baseado em aprendizagem (experiência) e disponibilizar este conhecimento para a aplicação em questão (ALECRIM, 2004).

3.1.1 Histórico

O trabalho pioneiro sobre a Teoria das Redes Neurais é atribuído a Warren McCulloch e Walter Pitts (o primeiro, médico, filósofo, matemático e poeta e o segundo, estatístico), de 1943, que descreveram um modelo ingenuamente simples de funcionamento de neurônio artificial. McCulloch e Pitts foram os primeiros a propor um modelo computacional para o neurônio biológico.

Em 1949, Hebb (neuropsicólogo canadense), propôs um postulado de aprendizagem baseado nas modificações sinápticas dos neurônios. Sua teoria foi posteriormente confirmada pela neurociência e exerceu influência considerável nas ciências da computação.

No fim da década de 50, Rosenblatt criou um modelo de aprendizagem supervisionado e chamou-o de *perceptron*, dando prosseguimento às idéias de McCulloch e Pitts. No modelo *perceptron*, o neurônio artificial é capaz de aprender segundo a idéia proposta por Hebb.

Nos anos 60, Widrow e Hoff utilizaram o *Adaline*, modelo similar ao *perceptron*, introduzindo o algoritmo da menor média quadrática.

Em 1969, Marvin Minsky e Seymour Pappert lançam o texto *Perceptrons*, expondo as limitações básicas de *perceptrons* isolados. A publicação de *Perceptrons* foi um divisor de águas que paralisou os estudos nesta área a partir da década de 70.

A fase obscura no estudo e desenvolvimento das Redes Neurais Artificiais encerrou-se apenas no início da década de 80 com o trabalho de

Hopfield, que propôs a implementação de memórias associativas por uma rede recorrente.

Em 1986, Rumelhart e McClelland impulsionaram novamente o uso do *Perceptron*, por meio de *Perceptrons* multicamadas e do algoritmo conhecido como *back-propagation*, o mais utilizado e difundido método até hoje para o treinamento da referida rede. Este algoritmo tornou viável a elaboração de redes mais complexas e robustas.

Segundo MASSAD (2004), os principais fatores que impulsionaram nas últimas três décadas a evolução das técnicas baseadas em redes neurais artificiais foram a disponibilização de recursos computacionais potentes e baratos, combinada com o desenvolvimento de algoritmos eficientes e de técnicas estatísticas baseadas em uso intensivo de computadores digitais.

3.1.2 Características Básicas de uma Rede Neural Artificial

Uma rede neural artificial é um processador maciçamente paralelo e distribuído, constituído de unidades de processamento simples (neurônio), que tem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso (HAYKIN, 2001).

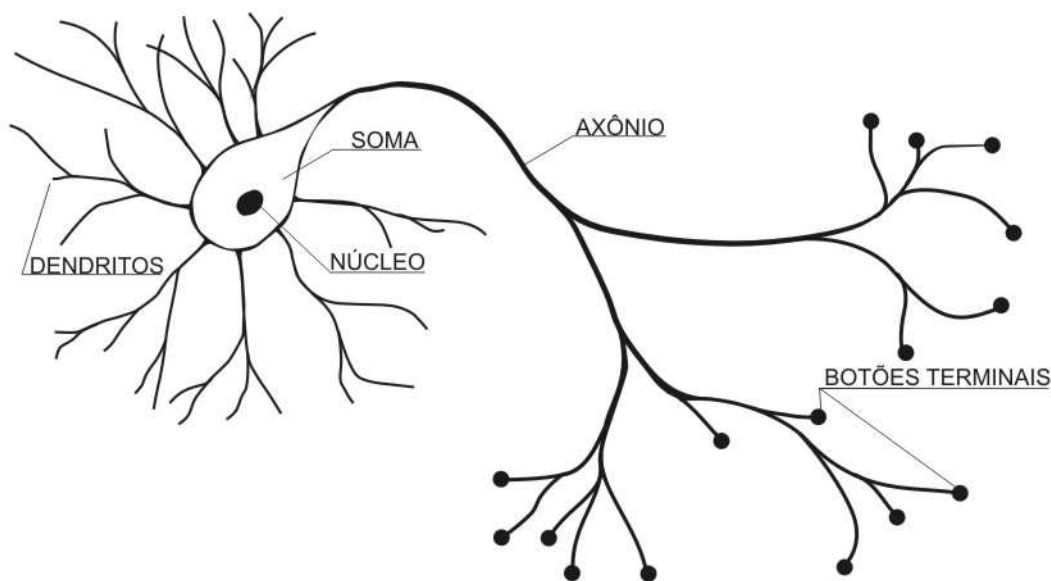
As unidades de processamento (neurônios artificiais) que compõem a RNA executam cálculos matemáticos, simulando, assim, o comportamento dos neurônios biológicos, com impressionante capacidade de processamento e armazenamento de informações.

O cérebro humano é constituído por cerca de 100 bilhões de neurônios. Cada neurônio processa e se comunica com outros milhares de neurônios continuamente e em paralelo.

Os movimentos e funções do nosso organismo estão relacionados ao funcionamento dos neurônios. Cada neurônio (Figura 3.1) consiste de um

corpo celular (soma), o qual contém o núcleo. A partir do corpo celular, existem prolongamentos chamados dendritos (que recebem o estímulo dos neurônios anteriores) e um prolongamento mais longo chamado de axônio (que estimula os neurônios subseqüentes).

FIGURA 3.1 REPRESENTAÇÃO BÁSICA DO NEURÔNIO BIOLÓGICO



FONTE: MEDEIROS, 2006

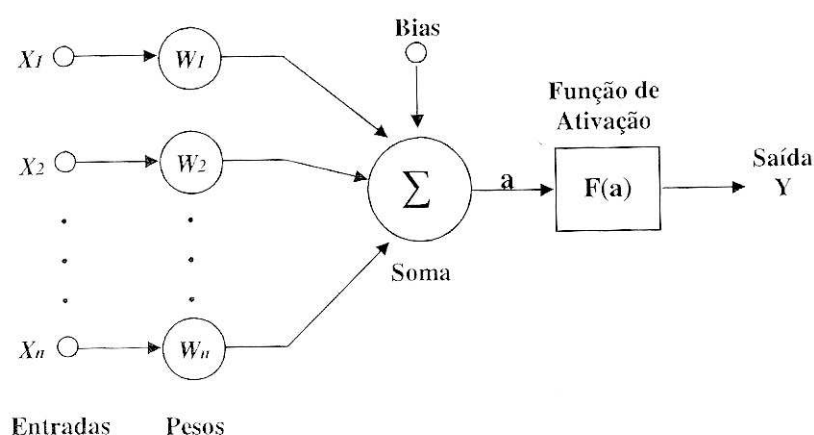
Os dendritos ligam-se ao redor da célula a outras células e o axônio faz uma conexão mais longa. A estas conexões dá-se o nome de sinapses (MEDEIROS, 2006).

É através das sinapses que os estímulos nervosos se propagam. Cada neurônio pode ser estimulado ou inibido por grande quantidade de neurônios, através de suas sinapses. A somatória dos estímulos e inibições determinará se o neurônio transmitirá o impulso ou não.

Uma RNA é composta por certo número de neurônios artificiais, conectados por ligações sinápticas ou *links*. Cada *link* possui uma quantidade associada como um peso. O conhecimento da rede é adquirido pela atualização dos pesos. Um neurônio artificial (Figura 3.2) possui um número n

de entradas x_1, x_2, \dots, x_n (dendritos). Cada sinal de entrada é multiplicado por um peso w_1, w_2, \dots, w_n , podendo ser negativo ou positivo. A seguir, a soma ponderada dos sinais de entrada aplicado ao neurônio, é executada. É aplicado o *bias* (θ), cuja função é aumentar o número de graus de liberdade do modelo, aumentando a capacidade da rede de ajustar-se ao conhecimento a ela fornecido. A soma é processada pela função de ativação para produzir a saída Y do neurônio (axônio).

FIGURA 3.2 EXEMPLO DE NEURÔNIO ARTIFICIAL



FONTE: MEDEIROS, 2006

3.1.3 Função de Ativação de uma Rede Neural

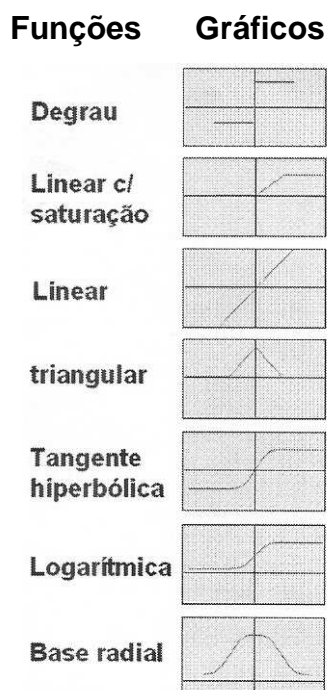
Segundo PALMA NETO (2005), o comportamento de um neurônio é determinado pelas funções a ele associadas e pelas conexões por meio das quais recebe entradas e transmite sua saída.

A função de ativação é muito importante para o comportamento de uma RN porque é ela que define a saída do neurônio artificial e, portanto, o caminho pelo qual a informação é conduzida (STEINER, 2006).

As funções mais utilizadas são as gaussianas, sigmóides, senoidais, lineares, tangentes hiperbólicas, logarítmicas, degrau, entre outras.

A Figura 3.3 mostra os gráficos das funções de ativação acima citadas.

FIGURA 3.3 FUNÇÕES DE ATIVAÇÃO



FONTE: LIMA, 2006

3.1.4 Treinamento de uma Rede Neural

O algoritmo de treinamento (ou aprendizado) de uma rede é um conjunto de regras por meio das quais os pesos das conexões são ajustados, usando, para o ajuste, um conjunto de exemplos que representa a função a ser aprendida. Esse conjunto de exemplos é geralmente chamado de conjunto de treinamento. Uma RNA sempre aprende a partir de um conjunto de treinamento; o aprendizado se reflete na capacidade de generalização que a rede vai exibir, quando usada em novas situações (PALMA NETO, 2005).

Segundo PALMA NETO (2005), os algoritmos de treinamento de RNAs pertencem a dois grupos de algoritmos: supervisionados e não-supervisionados.

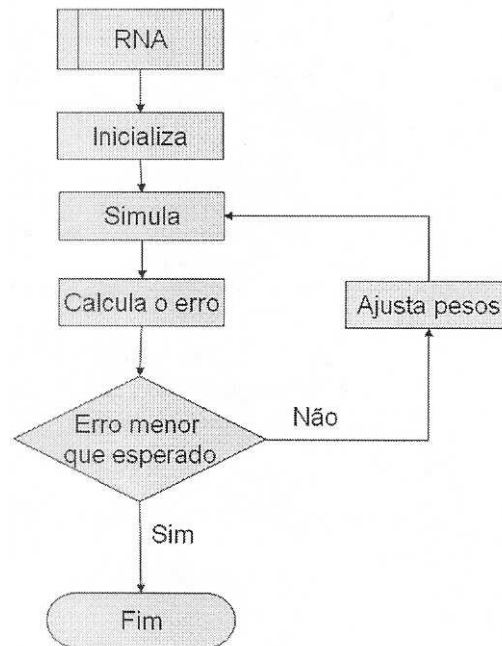
- a) No aprendizado supervisionado a descrição de cada exemplo de treinamento traz também a classe que o exemplo representa. Durante a fase de treinamento, o algoritmo tenta ajustar os pesos das conexões de maneira que a saída da rede coincida com a classe associada ao exemplo, para cada exemplo do conjunto de treinamento.
- b) No aprendizado não-supervisionado a rede é treinada sem considerar a informação da classe associada a cada exemplo de treinamento. O aprendizado não-supervisionado funciona por meio da modificação de pesos de conexões, de maneira que exemplos semelhantes entre si sejam associados à mesma unidade de saída.

O algoritmo padrão de uma RNA com aprendizado supervisionado segue os passos:

- a) A rede é inicializada aleatoriamente (os pesos são escolhidos de maneira aleatória);
- b) Um conjunto de dados (treinamento) é apresentado à rede e o erro entre o resultado da simulação e os dados reais é calculado;
- c) O erro propaga-se, ajustando os pesos de forma a minimizar o erro global;
- d) Uma nova simulação ocorre e o erro é calculado e propagado novamente;
- e) O processo repete-se até atingir um critério de parada.

A Figura 3.4 mostra o fluxograma do algoritmo genérico de uma RNA supervisionada.

FIGURA 3.4 ALGORITMO GENÉRICO DE UMA RNA SUPERVISIONADA



FONTE: LIMA, 2006

3.1.5 Fluxo de Dados em uma Rede Neural

A direção em que o fluxo de ativações acontece caracteriza as RNAs em:

- a) *feedforward* (acíclica): as conexões não formam ciclos. São rápidas para fornecer resposta a uma entrada. O *back-propagation* é um método numérico eficiente para o treinamento das redes *feedforward*.
- b) *feedback* (cíclica): as conexões podem formar ciclos. Cada vez que uma entrada é apresentada à rede, esta executa um processo iterativo antes de produzir a resposta. São redes mais difíceis de serem treinadas.

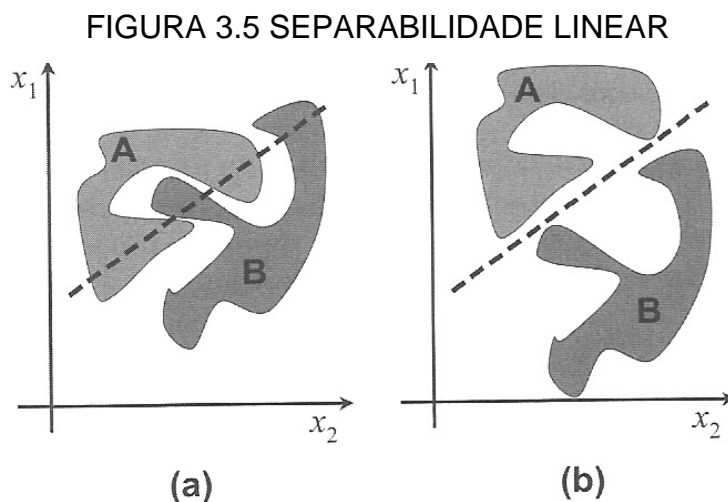
3.1.6 Modelos de Rede Neural

Nesta seção são apresentados os modelos básicos de RNAs: *Perceptron*, Redes Lineares e Redes de Múltiplas Camadas.

3.1.6.1 *Perceptron*

O *Perceptron* é creditado a Rosenblatt e constitui a forma mais simples de redes neurais artificiais. Foi amplamente estudado e usado em diversos trabalhos na década de 60 até cair em desuso após a constatação de que poderia apenas ser usado para a resolução de problemas linearmente separáveis.

A Figura 3.5 mostra a ilustração da separabilidade linear. Pode-se observar que na Figura 3.5(a) os subconjuntos não são linearmente separáveis, enquanto que na Figura 3.5(b) uma reta separa os dois subconjuntos A e B. Nos eixos, x_1 e x_2 são as variáveis independentes.



FONTE: MASSAD, 2004

O conceito de separabilidade linear pode ser estendido para o espaço tridimensional, em que o separador é um plano ou para o espaço n-

dimensional, em que duas ou mais categorias de dados são separáveis por um ou mais hiperplanos n -dimensionais (MASSAD, 2004).

Segundo STEINER (2006), o *Perceptron* pode ser visto como um instrumento de Reconhecimento de Padrões que não foi construído para reconhecer um conjunto específico de padrões, mas que tem alguma habilidade para aprender a reconhecer os padrões de um conjunto depois de um número finito de tentativas.

Os *perceptrons* são formados por uma única camada de neurônios e foram importantes para o desenvolvimento de outros tipos de rede (MASSAD, 2004).

3.1.6.2 Redes Lineares

Enquanto Rosenblatt trabalhava no perceptron, na mesma época, Widrow desenvolvia um modelo de rede neural linear, conceitualmente simples chamado ADALINE (ADaptive LINEar Element) e posteriormente, devido a sua generalização multidimensional, MADALINE (Múltipla ADALINE).

A relevância do ADALINE restringe-se ao contexto acadêmico. Widrow pôde contribuir realmente criando um princípio poderoso para as redes ADALINE “batizado” de Regra Delta ou Método do Gradiente. Sua proposta era obter o ponto de mínimo através de um processo de iteração local, usando apenas um exemplo de treinamento por vez.

Segundo KOVÁCS (2006), o método do gradiente aplicado ao problema de se determinar a melhor aproximação linear a um conjunto de dados, conhecido como Regra Delta, não passaria de um exercício acadêmico de otimização em espaços lineares, não tivesse preparado o terreno para a introdução, mais tarde, principalmente por Rumelhart, do mais poderoso procedimento que se conhece até hoje para o treinamento de RNAs, o algoritmo *back-propagation*.

3.1.6.3 Redes de Múltiplas Camadas

Segundo MASSAD (2004), para que uma rede neural tenha alguma habilidade em resolver problemas práticos, é necessário que ela represente funções lineares e não-lineares, assim como o algoritmo de treinamento deve ser suficientemente robusto e genérico, não restringindo a eficácia na resolução de problemas. Demonstrou-se que redes neurais de múltiplas camadas têm essas características.

O engenhoso algoritmo de retropropagação (*back-propagation*) foi pioneiramente proposto, em 1974, por Werbos e depois redescoberto e popularizado por Rumelhart e seus colaboradores em 1986.

Duas características das redes neurais de múltiplas camadas treinadas com o algoritmo *back-propagation*, que popularizou e difundiu seu uso nas mais diversas áreas científicas, são a simplicidade computacional para ajuste dos pesos sinápticos e a capacidade de representar relações não-lineares e lineares (MASSAD, 2004).

O algoritmo *back-propagation* decorre da minimização do erro quadrático entre as saídas da rede e as saídas desejadas, encontrando parâmetros que conduzam o erro quadrático assintoticamente para zero.

Segundo STEINER (2006), o número de neurônios ou camadas adequados não é de simples determinação. Pode-se usar diferentes conjuntos de soluções iniciais na procura de uma melhor solução para o problema.

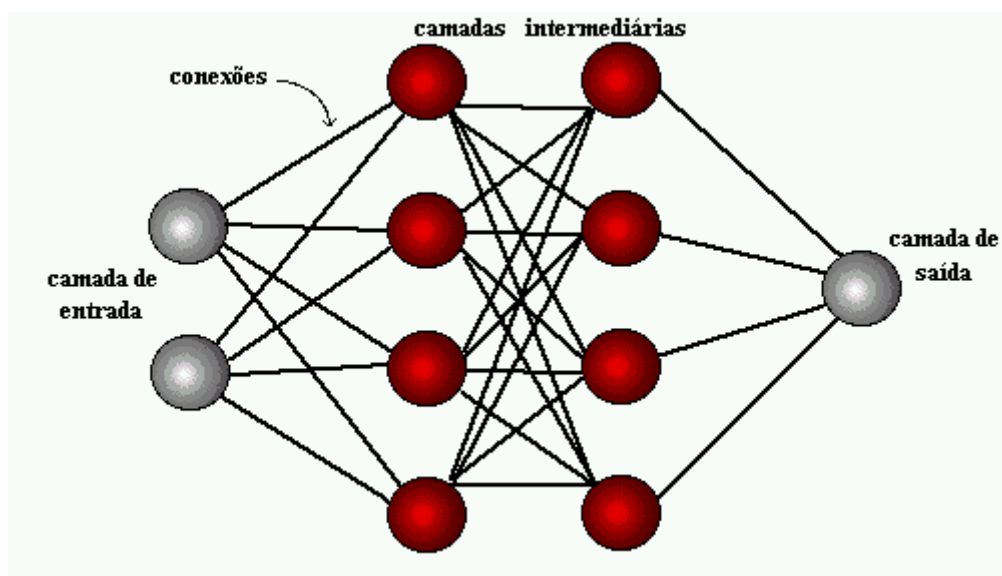
Uma rede *feedforward* com múltiplas camadas pode ser treinada com o algoritmo de retropropagação (*back-propagation*), adotando a função sigmóide que, a partir de entradas variando de $(-\infty, +\infty)$ gera saídas do tipo “0” ou “1”.

Arquiteturas neurais são organizadas em camadas classificadas em três grupos:

- a) Camada de entrada: onde os padrões (sinais de entrada) são apresentados à rede;
- b) Camadas ocultas ou escondidas: onde é feita a maior parte do processamento, através de conexões ponderadas, podendo ser consideradas extratoras de características. Estas camadas não têm contato com o ambiente externo à rede e têm o objetivo de melhorar o desempenho da rede, aumentando a possibilidade de divisão do espaço de entrada de forma não-linear;
- c) Camada de saída: onde o resultado final é concluído e apresentado.

A Figura 3.6 mostra a representação de uma rede neural com as camadas de entrada e de saída e duas camadas intermediárias.

FIGURA 3.6 ILUSTRAÇÃO DE UMA REDE DE MÚLTIPLAS CAMADAS



FONTE: CARVALHO, 2004.

A aprendizagem por retropropagação consiste de dois passos através das diferentes camadas da rede: um passo a frente, a propagação *forward*, que ativa a rede a partir das entradas e propaga para as saídas, e um passo a trás, a propagação *backward* ou retropropagação, que utiliza a saída desejada e a saída calculada pela rede para atualizar os pesos.

O processo de treinamento inicia-se com a apresentação do padrão de entrada para o algoritmo de retropropagação. Esse padrão é propagado por toda a rede até que uma saída seja gerada. O algoritmo de retropropagação utiliza, então, sua regra de generalização para determinar o erro produzido por cada unidade da rede. Finalmente, cada unidade modifica seu peso numa direção que reduza o sinal de erro e o processo é repetido para o próximo padrão.

Para uma RNA de três camadas: uma de entrada, uma oculta e uma de saída, o comportamento de cada unidade da rede pode ser modelado por funções matemáticas da seguinte forma: uma unidade i recebe os sinais de entrada da rede e agrupa-os de acordo com a equação (3.1).

$$i_i^p = \sum_{j=1}^n w_{ij} x_j^p + \theta_i, \quad p = 1, \dots, m \quad (3.1)$$

$$i = 1, \dots, k$$

$$j = 1, \dots, n$$

onde:

- i_i^p é a entrada da unidade i para o padrão p ;
- i é número de neurônios na camada oculta;;
- w_{ij} são as conexões peso entre as unidades i e j ;
- x_j^p é a conexão peso entre as unidades j para o padrão p ;
- θ_i são os *bias* das unidades i .

Com base no resultado encontrado em (3.1) é gerado um sinal de saída a_i^p , utilizando a função sigmoideal, como mostra a expressão (3.2).

$$a_i^p = \frac{1}{1 + e^{-i_i^p}} \quad (3.2)$$

Os sinais de saída são enviados para a única unidade de camada h , que os agrupa em i_h^p , conforme a função (3.3).

$$i_h^p = \sum_{i=1}^k w_{hi} a_i^p + \theta_h \quad (3.3)$$

onde:

i_h^p é a entrada da unidade h para o padrão p ;

i é número de neurônios na camada oculta;;

w_{hi} são as conexões peso entre as unidades h e i ;

a_i^p é o sinal de saída;

θ_h são os *bias* das unidades h .

A partir do resultado de (3.3) a saída a_h^p é gerada em (3.4).

$$a_h^p = \frac{1}{1 + e^{-i_h^p}} \quad (3.4)$$

Segundo STEINER (2006), é o vetor de pesos W que constitui o que a rede neural “sabe” e determina como ele responderá a qualquer entrada arbitrária do meio ambiente. Em geral, é muito difícil designar um W apropriado à tarefa de classificação. Uma solução geral é fazer com que a rede aprenda treinando-a com padrões.

O valor de saída obtido em (3.4) é comparado com o valor de saída desejado para o padrão p , d_p , calculando-se o erro quadrático, conforme (3.5).

$$E = \sum_{p=1}^{m+k} \frac{(d_p - a_h^p)^2}{2} \quad (3.5)$$

Finalizada a primeira fase (propagação *forward*), a segunda fase (propagação *backward*), que envolve as funções (3.6) a (3.10), mostradas a

seguir, executa um gradiente descendente em W objetivando minimizar o erro de (3.5).

A variação dos pesos w_{hi} , pode ser calculada segundo a função (3.6), a seguir.

$$\Delta_p w_{hi} = -\gamma \frac{\partial E_p}{\partial w_{hi}} \quad (3.6)$$

onde γ é a taxa de aprendizagem, $0 < \gamma < 1$.

Mas

$$\begin{aligned} \frac{\partial E_p}{\partial w_{hi}} &= -(d_p - a_h^p) \frac{e^{-i_h^p} a_i^p}{(1 + e^{-i_h^p})^2} \\ \frac{\partial E_p}{\partial w_{hi}} &= \frac{-1}{1 + e^{-i_h^p}} \frac{e^{-i_h^p}}{(1 + e^{-i_h^p})^2} [a_i^p (d_p - a_h^p)] \\ \frac{\partial E_p}{\partial w_{hi}} &= -a_h^p (1 - a_h^p) a_i^p (d_p - a_h^p) \end{aligned}$$

$$\text{Então, } \Delta_p w_{hi} = \gamma a_h^p (1 - a_h^p) a_i^p (d_p - a_h^p)$$

Levando em consideração a troca de pesos obtida no padrão $(t-1)$ na variação dos pesos para o padrão atual t , obtém-se a função (3.7).

$$\Delta_p w_{hi}(t) = \gamma a_h^p (1 - a_h^p) a_i^p (d_p - a_h^p) + \alpha \Delta_p w_{hi}(t-1) \quad (3.7)$$

onde α é a constante que determina o efeito na troca de pesos em $(t-1)$.

Os pesos entre as camadas oculta e de saída, em t , são determinados pela equação (3.8), a seguir.

$$w_{hi}(t) = w_{hi}(t-1) + \Delta_p w_{hi}(t) \quad (3.8)$$

A variação dos pesos w_{ij} , entre a camada escondida e a camada de entrada, pode ser calculada conforme os cálculos mostrados a seguir.

$$\Delta_p w_{ij} = -\gamma \frac{\partial E_p}{\partial w_{ij}}$$

$$\Delta_p w_{ij} = -(d_p - a_h^p) \frac{-(-e^{-i_h^p} \frac{\partial i_h^p}{\partial w_{ij}})}{(1 + e^{-i_h^p})^2}$$

Mas

$$\frac{\partial i_h^p}{\partial w_{ij}} = w_{hi} \frac{\partial a_h^p}{\partial w_{ij}}$$

$$\frac{\partial a_i^p}{\partial w_{ij}} = \frac{e^{-i_i^p} \frac{\partial i_h^p}{\partial w_{ij}}}{(1 + e^{-i_i^p})^2}$$

$$\frac{\partial i_i^p}{\partial w_{ij}} = x_j^p$$

Então,

$$\frac{\partial a_i^p}{\partial w_{ij}} = \frac{e^{-i_i^p} x_j^p}{(1 + e^{-i_i^p})^2} \text{ e } \frac{\partial i_h^p}{\partial w_{ij}} = w_{hi} \frac{e^{-i_i^p} x_j^p}{(1 + e^{-i_i^p})^2}$$

Assim,

$$\Delta_p w_{ij}(t) = \gamma (d_p - a_h^p) \frac{e^{-i_h^p}}{(1 + e^{-i_h^p})^2} w_{hi} \frac{e^{-i_i^p} x_j^p}{(1 + e^{-i_i^p})^2}$$

$$\Delta_p w_{ij}(t) = \gamma (d_p - a_h^p) a_h^p (1 - a_h^p) w_{hi} a_i^p (1 - a_i^p) x_j^p$$

Considerando na variação dos pesos para o padrão atual t a troca de pesos obtida para o padrão $(t-1)$, tem-se a equação (3.9).

$$\Delta_p w_{ij}(t) = \gamma (d_p - a_h^p) a_h^p (1 - a_h^p) w_{hi} a_i^p (1 - a_i^p) x_j^p + \alpha \Delta_p w_{ij}(t-1) \quad (3.9)$$

Desta maneira, os pesos entre as camadas de entrada e oculta, em t , podem ser obtidos segundo a equação (3.10), a seguir.

$$w_{ij}(t) = w_{ij}(t-1) + \Delta_p w_{ij}(t) \quad (3.10)$$

No ANEXO VI encontra-se a resolução de um exercício acadêmico de uma rede neural com duas camadas.

CAPÍTULO IV

4 IMPLEMENTAÇÃO DA REDE NEURAL ARTIFICIAL E ANÁLISE DOS RESULTADOS

4.1 IMPLEMENTAÇÃO DA REDE NEURAL ARTIFICIAL

Para a implementação das redes neurais do presente trabalho foram analisados e avaliados os diagnósticos de cefaléia de 2177 pacientes atendidos na Clínica Neurológica de Joinville / SC, no período compreendido entre 2002 e 2006. Todos os pacientes foram atendidos pelo mesmo médico, Ronald Moura Fiuza, neurologista e neurocirurgião que se dedica à cefaliatria. De cada diagnóstico foram extraídos os atributos que compõem o vetor de entradas da rede.

Os dados foram codificados, conforme a descrição já feita no capítulo II, objetivando aplicar valores numéricos para as variáveis qualitativas do questionário aplicado na anamnese de cada paciente.

As redes neurais artificiais do tipo múltiplas camadas foram criadas, treinadas e simuladas utilizando como ferramenta o software MATLAB 7.0 da *Mathworks* e o componente *Neural Networks Toolbox* utilizando o algoritmo *back-propagation*. O MATLAB permite a criação de diferentes tipos de redes e oferece uma liberdade bastante grande para a alteração dos parâmetros da rede.

Um algoritmo foi gerado no MATLAB para testar diferentes configurações de redes analisando, assim, a influência dos parâmetros nos resultados finais de simulação.

O número de camadas e o número de neurônios de cada camada devem ser definidos inicialmente.

Segundo HAYKIN (2001), apenas uma camada intermediária é necessária para aproximar qualquer função contínua. Assim, o presente trabalho é composto de uma camada de entrada, uma camada oculta e uma camada de saída.

As funções de ativação utilizadas neste trabalho foram:

- a função logística sigmoidal tanto na camada escondida quanto na camada de saída;
- a função tangente sigmoidal na camada escondida e a função linear na camada de saída;
- a função tangente sigmoidal na camada escondida e a função logística sigmoidal na camada de saída;
- a função logística sigmoidal na camada escondida e a função linear na camada de saída.

A rede foi inicializada com um número “x” de neurônios na camada escondida, treinada através de “y” épocas (ciclos; iterações), o erro foi calculado e o estado da rede salvo a cada 25 iterações. Este procedimento foi repetido “z” vezes. A seguir uma nova rede foi inicializada com um outro número de neurônios na camada escondida e assim sucessivamente. Por fim, o algoritmo encontrou a rede que apresentou o maior índice de acertos.

Desta forma, o algoritmo gerou “z” diferentes redes iniciais (com conjuntos de pesos diferentes, escolhidos aleatoriamente), com o mesmo número de neurônios na camada oculta, treinou-as por “y” iterações, e ao final do processo, apresentou aquela que mais se aproximou das saídas desejadas. A rede, então, foi testada para verificar a sua capacidade de generalização.

Os 2177 diagnósticos que compõem a amostra foram divididos em dois conjuntos. Para avaliar a generalização das redes treinadas, foi utilizado o

procedimento *holdout*, devido a grande base de dados. Este procedimento separa uma parte, (2/3) dos dados, para o treinamento da rede e outra, (1/3) dos dados, para testá-la. A amostragem foi estratificada, assegurando que cada classe estivesse representada de forma proporcional nos dois conjuntos de dados (treinamento e teste). A cada três pacientes, dois foram selecionados para o conjunto de treinamento e um para o conjunto de teste.

A partir do conjunto representado por 2/3 da amostra, realizou-se o treinamento supervisionado da rede. O outro conjunto, representado por 1/3 da amostra, chamado de conjunto de teste, é utilizado para verificar o aprendizado da rede.

A técnica proposta neste trabalho foi implementada da seguinte forma:

- Teste I: treinamento de uma RNA considerando 66 neurônios na camada de entrada (codificação 2), e cinco neurônios na camada de saída.
- Teste II: treinamento de cinco RNAs, cada uma delas com 66 neurônios na camada de entrada e um neurônio na camada de saída;
- Teste III: treinamento de uma RNA com 14 neurônios na camada de entrada (codificação 1) e cinco neurônios na camada de saída.

Os treinamentos das redes para os Testes I, II e III foram realizados variando os parâmetros definidos a seguir:

- número de neurônios na camada escondida: em cada teste, a rede foi treinada inicialmente sem a camada oculta e nos demais testes utilizando, respectivamente, 1, 2, 3, ..., 10 neurônios na camada oculta.
- conjuntos de pesos iniciais: em cada teste foram utilizados 5 conjuntos de pesos iniciais escolhidos aleatoriamente;
- funções de ativação: foram usadas 4 duplas de funções de ativação diferentes para cada teste, logística sigmoidal nas duas camadas;

logística sigmoidal e linear; logística sigmoidal e tangente sigmoidal; tangente sigmoidal e linear;

- número de neurônios na camada de entrada: para o Teste III o número de neurônios é igual a 14 (codificação 1) e para os Testes I e II o número é igual a 66 (codificação 2);
- número de neurônios na camada de saída: o número de neurônios na camada de saída é igual a cinco nos Testes I e III e igual a um no Teste II.

Segundo NIEVOLA (2006) uma decisão muito importante quando se está na etapa de treinamento de uma RNA é quando parar. Um dos critérios mais utilizados é o número de iterações.

No presente trabalho, o número de iterações foi fixado em 1000 ciclos. Em todas as iterações a taxa de aprendizagem utilizada foi $\gamma = 0,8$. Optou-se por não usar a taxa de *momentum* (α).

4.1.1 Teste I

No Teste I as redes foram treinadas utilizando a codificação 2 dos atributos de entrada. Desta forma, o vetor de entradas da rede possui 66 entradas binárias. A camada de saída é composta de 5 neurônios. O número de neurônios na camada escondida variou de “0” a “10”.

Os melhores resultados encontrados para cada topologia são mostrados na tabela 4.1, a seguir. Na primeira coluna tem-se o número de neurônios na camada oculta, na segunda coluna, a dupla de funções de ativação utilizada e na terceira coluna o percentual de acertos da rede.

TABELA 4.1 RESULTADOS ENCONTRADOS NO TESTE I

camada oculta	funções de ativação	acerto (%)
0	tansig, purelin	57,24
1	tansig, purelin	62,76
2	tansig, purelin	70,07
3	tansig, purelin	67,86
4	tansig, purelin	69,79
5	tansig, purelin	70,21
6	tansig, purelin	72,14
7	tansig, purelin	71,31
8	tansig, purelin	71,59
9	tansig, purelin	72,28
10	tansig, purelin	72,28
0	logsig, logsig	6,62
1	logsig, logsig	11,00
2	logsig, logsig	61,19
3	logsig, logsig	56,83
4	logsig, logsig	56,97
5	logsig, logsig	63,86
6	logsig, logsig	61,10
7	logsig, logsig	60,97
8	logsig, logsig	63,17
9	logsig, logsig	73,04
10	logsig, logsig	64,28
0	tansig, logsig	55,45
1	tansig, logsig	57,52
2	tansig, logsig	58,21
3	tansig, logsig	63,03
4	tansig, logsig	69,10
5	tansig, logsig	69,10
6	tansig, logsig	69,38
7	tansig, logsig	71,59
8	tansig, logsig	68,00
9	tansig, logsig	67,17
10	tansig, logsig	68,97
0	logsig, purelin	57,10
1	logsig, purelin	62,59
2	logsig, purelin	70,21
3	logsig, purelin	70,34
4	logsig, purelin	71,59
5	logsig, purelin	69,79
6	logsig, purelin	69,79
7	logsig, purelin	70,34
8	logsig, purelin	69,79
9	logsig, purelin	70,07
10	logsig, purelin	72,69

A partir da tabela 4.1 verifica-se que a melhor arquitetura é a que apresenta 9 neurônios na camada escondida usando a função de ativação logística sigmoidal tanto na camada escondida quanto na camada de saída. O percentual de acerto foi de 73,04%, o maior encontrado em todas as simulações realizadas.

4.1.2 Teste II

Neste Teste II foram treinadas cinco redes, cada uma delas com apenas uma saída. Cada uma das redes corresponde a um diagnóstico de cefaléia como saída, de acordo com a nomenclatura:

- RN I: Cefaléia Tensional;
- RN II: Enxaqueca sem Aura;
- RN III: Cefaléia por Uso Excessivo de Medicamentos;
- RN IV: Enxaqueca com Aura;
- RN V: Outros Tipos de Cefaléia.

O vetor de entradas é composto por 66 componentes binários, de acordo com a codificação 2 e a saída da rede é composta por apenas um neurônio. Os valores da saída estarão compreendidos no intervalo de (0, 1), sendo que os valores próximos a “1” ([0,5; 1)) representam determinado tipo de cefaléia (RN I, ..., RN V; um tipo para cada rede neural) e os valores próximos a “0” (0; 0,5) representam a não ocorrência daquele tipo de cefaléia. O número de neurônios na camada escondida variou de “0” a “10” para cada uma das redes.

As funções de ativação utilizadas para o Teste II foram a função logarítmica sigmoidal tanto na camada escondida quanto na camada de saída num primeiro momento. Em seguida, a função tangente sigmoidal na camada escondida e a função logarítmica sigmoidal na camada de saída foram utilizadas.

A tabela 4.2 mostra os melhores resultados obtidos para cada rede. A primeira coluna mostra o número de neurônios na camada oculta, a segunda coluna mostra as funções de ativação utilizadas, a terceira coluna mostra os percentuais de acerto para a RN I, a quarta coluna para RN II e assim sucessivamente, até a sétima coluna.

TABELA 4.2 RESULTADOS ENCONTRADOS NO TESTE II

camada oculta	funções de ativação	acerto (%) RN I	acerto (%) RN II	acerto (%) RN III	acerto (%) RN IV	acerto (%) RN V
0	logsig, logsig	71,38	87,86	94,90	95,45	84,41
1	logsig, logsig	83,59	92,69	95,45	98,07	85,38
2	logsig, logsig	83,17	92,28	95,45	97,93	85,24
3	logsig, logsig	82,62	92,41	95,45	97,75	85,66
4	logsig, logsig	83,59	92,41	95,45	98,07	86,21
5	logsig, logsig	82,21	91,59	95,45	98,30	85,66
6	logsig, logsig	83,17	92,41	95,45	97,52	85,93
7	logsig, logsig	82,21	91,86	95,72	97,93	85,66
8	logsig, logsig	82,07	92,14	95,59	98,53	86,48
9	logsig, logsig	82,90	92,14	95,45	98,48	86,07
10	logsig, logsig	82,62	92,14	95,45	97,93	85,66
0	tansig, logsig	71,66	85,93	95,03	95,45	84,28
1	tansig, logsig	82,07	92,55	95,45	95,45	85,24
2	tansig, logsig	82,34	94,99	95,45	95,45	85,66
3	tansig, logsig	82,21	92,69	95,45	95,45	85,66
4	tansig, logsig	82,62	92,69	95,45	95,45	85,52
5	tansig, logsig	83,45	92,69	95,59	95,45	85,79
6	tansig, logsig	82,76	92,97	95,45	95,45	85,66
7	tansig, logsig	82,90	92,83	95,73	95,45	85,66
8	tansig, logsig	83,03	92,69	95,59	95,45	88,24
9	tansig, logsig	81,66	92,69	95,59	95,45	85,52
10	tansig, logsig	87,83	92,97	95,59	95,45	85,66

A tabela 4.3, a seguir, mostra os melhores resultados obtidos para as cinco redes neurais treinadas, a partir dos resultados visualizados na tabela 4.2. A primeira coluna representa a rede treinada, a segunda coluna mostra as funções que melhor desempenharam para cada rede, a terceira coluna traz o número de neurônios da camada escondida e a última coluna mostra o valor percentual de acerto de cada rede treinada.

TABELA 4.3 MELHORES RESULTADOS TESTE II

REDE NEURAL	FUNÇÕES DE ATIVAÇÃO	CAMADA OCULTA	PERCENTUAL DE ACERTO
RN I	tansig, logsig	10	87,83
RN II	tansig, logsig	2	94,99
RN III	tansig, logsig	7	95,73
RN IV	logsig, logsig	8	98,53
RN V	tansig, logsig	8	88,24

4.1.3 Teste III

No Teste III as redes foram treinadas utilizando a codificação 1 dos atributos de entrada. Desta forma, o vetor de entradas da rede possui 14 entradas. A camada de saída é composta de 5 neurônios. O número de neurônios na camada escondida variou de “0” a “10”.

A tabela 4.4 traz os melhores resultados encontrados para cada topologia. Na primeira coluna verifica-se o número de neurônios na camada oculta; na segunda coluna, a dupla de funções de ativação utilizada e na terceira coluna o percentual de acertos da rede.

Foram usadas no Teste III as funções de ativação logarítmica sigmoidal tanto na camada escondida quanto na camada de saída, e, em seguida, a função tangente sigmoidal na camada escondida e a função logarítmica sigmoidal na camada de saída.

TABELA 4.4 RESULTADOS ENCONTRADOS NO TESTE III

camada oculta	funções de ativação	acerto (%)
0	logsig, logsig	0,00
1	logsig, logsig	0,00
2	logsig, logsig	54,57
3	logsig, logsig	53,13
4	logsig, logsig	49,52
5	logsig, logsig	45,34
6	logsig, logsig	56,82
7	logsig, logsig	54,25
8	logsig, logsig	56,55
9	logsig, logsig	63,16
10	logsig, logsig	63,67
0	tansig, logsig	0,00
1	tansig, logsig	56,18
2	tansig, logsig	57,46
3	tansig, logsig	57,60
4	tansig, logsig	55,81
5	tansig, logsig	54,98
6	tansig, logsig	57,05
7	tansig, logsig	63,94
8	tansig, logsig	59,94
9	tansig, logsig	63,16
10	tansig, logsig	66,65

A tabela 4.4 mostra que a melhor arquitetura é a que apresenta 10 neurônios na camada escondida usando a função de ativação tangente sigmoidal na camada escondida e a função de ativação logarítmica sigmoidal na camada de saída. O percentual de acerto foi de 66,65%, o maior encontrado em todas as iterações realizadas.

4.2 ANÁLISE DOS RESULTADOS

Na seção 4.1, deste capítulo, foram apresentados os treinamentos de algumas RNAs e suas diferentes topologias, com o objetivo de obter um diagnóstico preditivo para os pacientes com cefaléia atendidos pelo médico Ronald Moura Fiuza.

Os melhores resultados foram obtidos a partir da codificação 2 das variáveis de entrada, processo que transformou-as em vetores de coordenadas binárias.

A tabela 4.5, a seguir, mostra os melhores resultados de todos os treinamentos realizados, especificando a topologia, o número de neurônios na camada de entrada, o número de neurônios na camada escondida, o número de neurônios na camada de saída, a tipo de codificação das variáveis de entrada e o percentual de acerto da rede em cada simulação.

TABELA 4.5 MELHORES RESULTADOS DAS REDES NEURAIS

TESTE	NEURÔNIOS			CODIFICAÇÃO	FUNÇÕES DE ATIVAÇÃO	PERCENTUAL DE ACERTO
	ENTRADA	OCULTA	SAÍDA			
I	66	9	5	2	LOGSIG / LOGSIG	73,04
II	66	10	1 - RN I	2	TANSIG / LOGSIG	87,83
II	66	2	1 - RN II	2	TANSIG / LOGSIG	94,99
II	66	7	1 - RN III	2	TANSIG / LOGSIG	95,73
II	66	8	1 - RN IV	2	LOGSIG / LOGSIG	98,53
II	66	8	1 - RN V	2	TANSIG / LOGSIG	88,24
III	14	10	5	1	TANSIG / LOGSIG	66,65

O Teste II apresentou os melhores resultados para a predição do diagnóstico de pacientes com cefaléia, como mostrou a tabela 4.5. Constatase que cinco RNAs, com apenas uma saída cada, têm desempenho melhor do que uma RNA com cinco saídas e que a codificação das entradas e saídas em coordenadas binárias facilita o processamento da rede.

Segundo MARQUES (2006), para avaliar a classificação de uma ferramenta de Reconhecimento de Padrões pode-se calcular a Taxa Aparente de Erro (*APER*), definida como sendo a fração das observações no treinamento amostral referente ao reconhecimento errôneo pela função obtida. Esta taxa é calculada pela Matriz de Confusão, que apresenta a real situação das observações nos grupos comparando-a com o reconhecimento apresentado pelo modelo encontrado. Para n_1 observações de π_1 e n_2 observações de π_2 , a matriz de confusão tem a forma dada pelo quadro 4.1, mostrado a seguir.

QUADRO 4.1 MATRIZ DE CONFUSÃO

		classificação prevista	
		π_1	π_2
classificação real	π_1	$n_{1/1}$	$n_{1/2}$
	π_2	$n_{2/1}$	$n_{2/2}$

Onde: n_{ij} são as observações de π_i classificadas corretamente (se $i = j$) ou incorretamente (se $i \neq j$).

Assim:

- $n_{1/1}$ é o número de observações de π_1 classificados corretamente como de π_1 ;
- $n_{2/1}$ é o número de observações de π_2 classificados incorretamente como de π_1 ;
- $n_{2/2}$ é o número de observações de π_2 classificados corretamente como de π_2 ;
- $n_{1/2}$ é o número de observações de π_1 classificados incorretamente como de π_2 .

A Taxa Aparente de Erro é calculada pela fórmula mostrada a seguir (equação 4.1).

$$APER = \frac{n_{1/2} + n_{2/1}}{n_1 + n_2} \quad (4.1)$$

Na equação 4.1, n_1 representa o tamanho da amostra 1 e n_2 , da amostra 2. A *APER* deve ser interpretada como a proporção de observações classificadas incorretamente.

4.2.1 Teste I

A partir do melhor resultado obtido no treinamento da rede neural do Teste I, pode-se visualizar a Matriz de Confusão, no quadro 4.2, a seguir.

QUADRO 4.2 MATRIZ DE CONFUSÃO – TESTE I

		CLASSIFICAÇÃO PREVISTA					
		CT	ECA	ESA	CEM	OUT	S/C
CLASSIFICAÇÃO REAL	CT	1125	0	13	0	35	79
	ECA	4	44	18	0	7	26
	ESA	13	1	250	0	0	43
	CEM	65	3	12	0	1	19
	OUT	131	1	30	0	171	86

Legenda:

CT – Cefaléia Tensional

ECA – Enxaqueca com Aura

ESA – Enxaqueca sem Aura

CEM – Cefaléia por Uso Excessivo de Medicamentos

OUT – Outros Diagnósticos de Cefaléia

S/C – Sem Classificação

A Taxa Aparente de Erro para o Teste I é igual a $0,2696$ ($13 + 35 + 79 + 44 + 18 + 7 + 26 + 13 + 1 + 43 + 65 + 3 + 12 + 1 + 19 + 131 + 1 + 30 + 86$) / 2177 , ou seja, $26,96\%$ dos diagnósticos foram classificadas incorretamente para a melhor simulação.

4.2.2 Teste II

O Teste II representa o treinamento de cinco redes: RN I, RN II, RN III, RN IV E RN V.

Para cada rede foi construída a Matriz de Confusão, segundo os quadros 4.3 a 4.7, mostrados a seguir.

QUADRO 4.3 MATRIZ DE CONFUSÃO – RN I – TESTE II

		classificação prevista	
		CT	ODQ
classificação real	CT	1156	96
	ODQ	169	756

Legenda:

ODQ – Outro Diagnóstico Qualquer

QUADRO 4.4 MATRIZ DE CONFUSÃO – RN II – TESTE II

		classificação prevista	
		ESA	ODQ
classificação real	ESA	306	1
	ODQ	108	1762

QUADRO 4.5 MATRIZ DE CONFUSÃO – RN III – TESTE II

		classificação prevista	
		CEM	ODQ
classificação real	CEM	100	0
	ODQ	93	1984

QUADRO 4.6 MATRIZ DE CONFUSÃO – RN IV – TESTE II

		classificação prevista	
		ECA	ODQ
classificação real	ECA	99	0
	ODQ	32	2046

QUADRO 4.7 MATRIZ DE CONFUSÃO – RN V – TESTE II

		classificação prevista	
		OUT	ODQ
classificação real	OUT	399	20
	ODQ	236	1522

A tabela 4.6, a seguir, mostra o valor da Taxa Aparente de Erro para cada rede treinada no Teste II, a seguir.

TABELA 4.6 TAXA APARENTE DE ERRO

REDE NEURAL	APER
RN I	0,1217
RN II	0,0501
RN III	0,0427
RN IV	0,0147
RN V	0,1176

Assim, para a RN I, o erro é de 12,17%; para RN II, o erro é de 5,01%; para a RN III, o erro é de 4,27%; para a RN IV, o erro é de 1,47%; e para a RN V, o erro é de 11,76%.

4.2.3 Teste III

Pode-se visualizar a Matriz de Confusão para o Teste III no quadro 4.8, calculada a partir do melhor resultado obtido para o treinamento da rede segundo a topologia mostrada na tabela 4.4.

QUADRO 4.8 MATRIZ DE CONFUSÃO – TESTE III

		CLASSIFICAÇÃO PREVISTA					
		CT	ECA	ESA	CEM	OUT	S/C
CLASSIFICAÇÃO REAL	CT	1099	0	8	0	38	107
	ECA	13	32	24	0	4	26
	ESA	39	11	195	0	18	44
	CEM	50	1	9	22	1	17
	OUT	86	2	47	0	103	181

O valor da Taxa Aparente de Erro é igual a 0,3335, ou seja, 33,35% dos diagnósticos foram classificados erroneamente para a melhor simulação.

Assim, as RNAs aqui apresentadas e trabalhadas, poderão auxiliar o médico especialista (cefaliatra) no diagnóstico preditivo dos tipos mais freqüentes de cefaléia, porém, nunca serão usadas como uma ferramenta substitutiva pois a figura do médico é indispensável.

CAPÍTULO V

5 CONSIDERAÇÕES FINAIS

5.1 CONCLUSÕES

As RNAs constituem uma técnica conhecida por simular o comportamento do cérebro humano, utilizando a tentativa e erro no processo de acumulação de novos conhecimentos.

Sua utilização é ampla e abrangente, sendo reconhecida sua utilidade no processamento de informações derivadas dos setores comerciais, industriais e de serviços, públicos e privados. Há exemplos de utilização com sucesso na área da saúde, onde podem se mostrar instrumentos de valor para médicos, administradores e demais profissionais de saúde.

No presente trabalho, as RNAs foram exploradas para resolver o problema de predição dos diagnósticos mais freqüentes das dores de cabeça, que constituem um dos mais freqüentes sintomas que afligem a população e que acionam o sistema de saúde em todo o mundo.

Utilizou-se o banco de dados de uma clínica especializada, que dispõe em seu sistema de questionário semi-estruturado, cujo preenchimento resultou nos dados que municiaram as entradas das RNAs.

Todas as RNAs foram treinadas, separadamente, a partir do aprendizado supervisionado. Foi realizada, empiricamente, a variação dos seguintes parâmetros, com o objetivo de encontrar a melhor estrutura de rede: funções de ativação das camadas de entrada e oculta; número de neurônios nas camadas de entrada e saída; número de neurônios na camada oculta.

Utilizou-se o algoritmo *back-propagation*, baseado na generalização do mapeamento de um padrão de entrada para um espaço de saída através da minimização do erro entre a saída que ele produziu e o padrão de saída fornecido.

A utilização de RNAs apresentou-se adequada no tratamento das informações dos 2177 pacientes. A base de dados foi implementada a partir dos dados (sintomas) de cada paciente e do conhecimento do especialista (diagnóstico).

Verificou-se que o modelo que apresentou um melhor desempenho utilizou 5 redes neurais com uma saída, onde os percentuais de acerto variaram de 87,83% a 98,53%.

As RNAs mostraram-se uma ferramenta de grande valia para o Reconhecimento de Padrões no diagnóstico médico dos tipos mais freqüentes de cefaléia. Desta forma, utilizando as informações obtidas no questionário, os médicos têm condições de fazer um diagnóstico preditivo do paciente com dor de cabeça, fazendo uso da ferramenta.

Predizendo os diagnósticos mais freqüentes de cefaléia, o presente instrumento pode ser utilizado como triagem, otimizando os recursos disponíveis no sistema de saúde. Assim, na porta de entrada do sistema, um agente de saúde ou um enfermeiro pode ser treinado para a aplicação do questionário e a obtenção da hipótese diagnóstica. Com a adequada supervisão, pode ser desenhado um quadro capaz de orientar o paciente neste ambiente de atendimento primário. Os casos mais simples podem ser encaminhados para o acompanhamento de médico de família. Os casos mais complexos podem ser encaminhados ao especialista e mesmo ao hospital. Muitas consultas e exames desnecessários podem ser evitados, desafogando a rede de saúde. Em última análise o modelo pode propiciar agilidade, economia e funcionalidade ao sistema, no que se refere aos numerosos pacientes com cefaléia.

5.2 SUGESTÕES PARA TRABALHOS FUTUROS

As sugestões, a partir dos resultados do presente trabalho, para o desenvolvimento de futuros trabalhos são:

- a utilização de RNAs para o mesmo banco de dados variando a taxa de aprendizagem, incluindo a taxa de *momentum* e trabalhando com uma variação maior na quantidade de neurônios na camada oculta;
- o emprego de outras técnicas de Reconhecimento de Padrões para a resolução do problema como, por exemplo, métodos estatísticos, dentre os quais, pode-se citar: Análise Discriminante, Regressão Logística; assim como o de outras meta-heurísticas (além de Redes Neurais) como, por exemplo, Algoritmos Genéticos, sempre objetivando a comparação dos resultados entre as diversas ferramentas, procurando pelo método que apresente um melhor desempenho;
- a resolução do problema usando Sistemas Especialistas. Os Sistemas Especialistas são sistemas que aplicam mecanismos automatizados de raciocínio e são concebidos para reproduzir o comportamento de especialistas humanos altamente experientes;
- a separação dos diagnósticos em mais do que cinco grupos distintos, já que a Sociedade Internacional de Cefaléia descreve com pormenores mais de 200 doenças capazes de causar dores de cabeça.

REFERÊNCIAS

ADAMOWICZ, E. C. *Reconhecimento de Padrões na Análise Econômico-Financeira de Empresas*. Universidade Federal do Paraná, Curitiba, PR, 2000. (Dissertação de Mestrado)

ALECRIM, E. *Redes Neurais Artificiais*. Disponível em: <http://www.infowester.com/redesneurais.php>. Acesso em: 14 de agosto de 2007.

BALDIN, M. S. Z. *Redes Neurais na Orientação Profissional*. Universidade Federal do Paraná, Curitiba, PR, 2002. (Dissertação de Mestrado)

BAPTISTELLA, M. *O uso de Redes Neurais e Regressão Linear Múltipla na Engenharia de Avaliações: Determinação dos Valores Venais de Imóveis Urbanos*. Universidade Federal do Paraná, Curitiba, PR, 2005. (Dissertação de Mestrado)

BAREA, LM; FORCELINI, CM. *A epidemiologia das cefaléias primárias*. Lemos Editorial: São Paulo, 2002.

BENSEÑOR et al. *Validation of a questionnaire for the diagnosis of headache in an outpatient clinic at a university hospital*. Arq neuropsiquiatr. 1997; 55: 364-9.

BERG, J; RAMADAN; N.M. *Societal burden of the headache*. The Headaches. 3rd Edition. Williams and Wilkins. Philadelphia, 2006.

CARVALHO, A. P. L. *Redes Neurais Artificiais*. Disponível em: <http://www.icmc.usp.br/~andre/research/neural/index.htm#links>. São Paulo, SP, 2004. Acesso em: 10 fev. 2007.

Classificação Internacional das Cefaléias. 2ª. Edição, São Paulo: Editora Segmento Farma, 2004.

DAHLÖF, C. G. H; SOLOMON, G. D. *Impact of headache on the individual and family*. The Headaches. 3rd Edition. Williams and Wilkins. Philadelphia, 2006.

HANSELMAN, D; LITTLEFIELD, B. *MATLAB 6: Curso Completo*. São Paulo: Prentice Hall, 2003.

HAYKIN, S. *Redes neurais*. Porto Alegre: Bookman, 2001.

KOVÁCS, Z. L. *Redes Neurais Artificiais: Fundamentos e Aplicações*. 4ª. Edição. São Paulo: Editora Livraria da Física, 2006.

LE MOS, E. P. *Análise de Crédito Bancário com o Uso de Data Mining: Redes Neurais e Árvores de Decisão*. Universidade Federal do Paraná, Curitiba, PR, 2003. (Dissertação de Mestrado)

LIMA, J. P. Hl. *Redes neurais artificiais aplicadas à otimização de processos de deposição de filmes finos poliméricos*. São Paulo, Universidade de São Paulo, São Paulo, SP, 2006. Disponível em: <http://www.teses.usp.br/teses/disponiveis/3/3140/tde-05092006-130151/> (Dissertação de Mestrado). Acesso em: 10 fev. 2007.

LIPTON, R. B; BIGAL, M. E. *Ten lessons on epidemiology of migraine*. Headache, 2007.

MARANHÃO FILHO, P. *História das Cefaléias*. Lemos Editorial: São Paulo, SP, 2002.

MARQUES, J. M. *Análise Discriminante*. Notas de aula apresentadas na disciplina Análise Multivariada Aplicada à Pesquisa, Universidade Federal do Paraná, Curitiba, PR, 2006.

MARTINS, L. W; ASSIS, J. T; MONAT, A S. *Aplicação de Redes Neurais para o Diagnóstico Diferencial da Doença Meningocócica*. IX Congresso Brasileiro em Informática em Saúde, Ribeirão Preto, SP, 2004. Disponível em: <http://www.sbis.org.br/cbis9/arquivos/627.PDF>.

MASSAD, E; MENESES, R. X.; SILVEIRA, P. S. S; ORTEGA, N. R. S. *Métodos Quantitativos em Medicina*. Barueri, SP: Editora Manole, 2004.

MEDEIROS, L. F. *Redes Neurais em Delphi*. 2ª. Edição. Florianópolis: Editora Visual Books, 2006.

NIEVOLA, Júlio César. *Redes Neurais Artificiais*. Disponível em: http://www.ppgia.pucpr.br/~euclidesfjr/Metodos_Inteligentes/0206/RNA-XIIERI.pdf. Acesso em: 10 fev. 2007.

OLESEN, J; DODICK, D. W. *The history and examination of headache patients*. The Headaches. 3rd Edition. Williams and Wilkins. Philadelphia, 2006.

Organização Mundial de Saúde: CID 10. Décima revisão. Edusp. São Paulo, SP, 1996.

PALMA NETO, L. G; NICOLETTI, M. C. *Introdução às Redes Neurais Construtivas*. São Carlos, SP: Editora da Universidade Federal de São Carlos, 2005.

PAVANELLI, G. *Análise do tempo de Duração de Processos Trabalhistas Utilizando Redes Neurais Artificiais como Apoio à Tomada de Decisões*.

Universidade Federal do Paraná, Curitiba, PR, 2007. (Dissertação de Mestrado)

PRECHELT, L. *PROBEN 1 – A Set of Neural Network Benchmark Problems and Benchmarking Rules*. Universidade Karlsruhe. Karlsruhe, Alemanha, 1994.

RASSMUNSEN, B. K; JENSEN, R; OLESEN, J. *Questionnaire versus clinical interview in the diagnosis of headache*. Headache 1991; 31: 290-5

SANTOS, A. M; PEREIRA, B. B; SEIXAS, J. M; MELLO, F. C; KRITSKI, A. Redes Neurais Artificiais Aplicadas ao Diagnóstico de Tuberculose Pulmonar Paucibacilar. XI Simpósio de Estatística Aplicada à Experimentação Agrônômica, Londrina, PR, 2005. Disponível em: <http://www.posgraduacao.ufla.br/gauss/congresso/11seagro/conteudo/arquivos/219.pdf>.

STEINER, M. T. A. *Redes Neurais*. Notas de aula apresentadas na disciplina Aplicações de Pesquisa Operacional, Universidade Federal do Paraná, 2006.

STEINER, M. T. A. *Uma Metodologia para o Reconhecimento de Padrões Multivariados com Resposta Dicotômica*. Curitiba, Universidade Federal de Santa Catarina, 1995. (Tese de Doutorado)

STERGIOUS, C; SIGANOS, D. *Neural Networks*. Disponível em: http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html. Acesso em: 20 mar. 2007.

STOVNER, L. J; SCHER, A. I. *Epidemiology of headache*. The Headaches. 3rd Edition. Williams and Wilkins. Philadelphia, 2006.

TATIBANA, C. Y; KAETSU, D. Y. *Uma Introdução às Redes Neurais*. Disponível em: <http://www.din.uem.br/ia/neurais/>. Acesso em: 13 dez. 2006.

VIEIRA, S. R. T; PINHEIRO NETO, H. *Um Mecanismo de Recuperação de Informação na WEB baseado em Redes Neurais*. Disponível em: <http://www.cientefico.frb.br/2005.2/BCC/CC.NETO&.F3.pdf>

ANEXOS

ANEXO I

QUESTIONÁRIO (ANAMNESE DO PACIENTE)

- | | | |
|---|--|--|
| <p>1) Sexo</p> <p><input type="checkbox"/> Feminino</p> <p><input type="checkbox"/> Masculino</p> | <p>7) Surgimento da Dor</p> <p><input type="checkbox"/> Crise</p> <p><input type="checkbox"/> Contínua</p> <p><input type="checkbox"/> Episódica</p> <p><input type="checkbox"/> Salvas</p> | <p><input type="checkbox"/> Fotofobia</p> <p><input type="checkbox"/> Fonofobia</p> <p><input type="checkbox"/> Transtornos Visuais</p> <p><input type="checkbox"/> Tontura</p> <p><input type="checkbox"/> Febre</p> |
| <p>2) Idade</p> <p><input type="checkbox"/> Até 12 anos</p> <p><input type="checkbox"/> 13 a 19 anos</p> <p><input type="checkbox"/> 20 a 39 anos</p> <p><input type="checkbox"/> 40 anos ou mais</p> | <p>8) Evolução da Dor</p> <p><input type="checkbox"/> Estável</p> <p><input type="checkbox"/> Mudando</p> <p><input type="checkbox"/> Progressiva</p> | <p>12) O que causa ou piora a dor</p> <p><input type="checkbox"/> Nervosismo</p> <p><input type="checkbox"/> Menstruação</p> <p><input type="checkbox"/> Alguns Alimentos</p> <p><input type="checkbox"/> Bebidas</p> <p><input type="checkbox"/> Sono Irregular</p> <p><input type="checkbox"/> Esforço Físico</p> <p><input type="checkbox"/> Sexo</p> |
| <p>3) Início da dor</p> <p><input type="checkbox"/> Dias</p> <p><input type="checkbox"/> Semanas</p> <p><input type="checkbox"/> Meses</p> <p><input type="checkbox"/> Anos</p> | <p>9) Frequência da Dor</p> <p><input type="checkbox"/> Diária</p> <p><input type="checkbox"/> Mais de 15 episódios ao mês</p> <p><input type="checkbox"/> 1 a 2 episódios por semana</p> <p><input type="checkbox"/> 1 a 3 episódios ao mês</p> <p><input type="checkbox"/> alguns episódios ao ano</p> | <p>13) O que melhora</p> <p><input type="checkbox"/> Deitar</p> <p><input type="checkbox"/> Relaxar</p> <p><input type="checkbox"/> Distrair</p> <p><input type="checkbox"/> Andar</p> <p><input type="checkbox"/> Gravidez</p> |
| <p>4) Localização da Dor</p> <p><input type="checkbox"/> Hemicrânea</p> <p><input type="checkbox"/> Holocraniana</p> <p><input type="checkbox"/> Frontal / Face</p> <p><input type="checkbox"/> Occipital / Nuca</p> <p><input type="checkbox"/> Localizada</p> | <p>10) Duração do Episódio</p> <p><input type="checkbox"/> segundos</p> <p><input type="checkbox"/> minutos</p> <p><input type="checkbox"/> horas</p> <p><input type="checkbox"/> dias</p> <p><input type="checkbox"/> semanas</p> | <p>14) Usa analgésicos?</p> <p><input type="checkbox"/> Nunca</p> <p><input type="checkbox"/> Raramente</p> <p><input type="checkbox"/> Sempre que dói</p> <p><input type="checkbox"/> quase todos os dias</p> <p><input type="checkbox"/> Diariamente</p> <p><input type="checkbox"/> Vários ao dia</p> |
| <p>5) Intensidade da Dor</p> <p><input type="checkbox"/> Forte</p> <p><input type="checkbox"/> Muito Forte</p> <p><input type="checkbox"/> Moderada</p> <p><input type="checkbox"/> Fraca</p> | <p>11) Sintomas Associados a Dor</p> <p><input type="checkbox"/> Náusea</p> <p><input type="checkbox"/> Vômito</p> | |
| <p>6) Tipo de Dor</p> <p><input type="checkbox"/> Latejante</p> <p><input type="checkbox"/> Peso</p> <p><input type="checkbox"/> Pontada</p> <p><input type="checkbox"/> Atípica</p> | | |

ANEXO II

DADOS ORIGINAIS

Sexo	Data Nascimento	CID		Início	Localização	Intensidade	Tipo
M	25/05/1949	G44.2	10	Anos	Frontal/Face	Forte	Latejante
M	25/05/1949					Moderada	
F	27/05/1946	G44.2	2	Meses	Hemocrânia	Forte	Atípica
F	27/05/1946				Frontal/Face		
M	11/03/1978	G44.2	1	Meses	Holocrânia	Forte	Latejante
M	20/09/1973	G44.2	1	Anos	Holocrânia	Moderada	Latejante
M	21/11/1936	G44.2	2	Semanas	Occipital/Nuca	Moderada	Atípica Peso ou Pressão
F	26/09/1970	G44.2	5	Anos	Frontal/Face	Forte	
F	26/09/1970					Moderada	
F	24/07/1937	F34.1	2	Meses	Holocrânia	Moderada	Latejante
F	24/07/1937	G44.2					
M	14/02/1954	F41.9	3	Meses	Occipital/Nuca	Moderada	Latejante
M	14/02/1954	G44.2					
F	01/04/1961	R51 [1	Semanas	Frontal/Face	Forte	
F	01/04/1961				Occipital/Nuca		
M	10/06/1957	G44.4	2	Anos		Forte	Peso ou Pressão
M	10/06/1957					Moderada	Atípica
M	17/09/1971	G43.9	5	Anos	Hemocrânia	Forte	Latejante
M	17/09/1971				Holocrânia	Moderada	Peso ou Pressão
M	09/10/1964	G43.9	20	Anos	Frontal/Face	Moderada	Latejante
M	09/10/1964						Pontadas
M	12/05/1965	R51 [10	Dias	Holocrânia	Fraca	Peso ou Pressão
M	10/04/1964	G43.1	20	Anos	Hemocrânia	Forte	
M	10/04/1964	G44.2			Frontal/Face		
M	10/04/1964						
M	28/05/1984	F06.7	3	Anos	Holocrânia	Forte	Pontadas
M	28/05/1984	G44.2				Moderada	
F	15/01/1974	F34.1	4	Anos	Holocrânia	Forte	Peso ou Pressão
F	15/01/1974	G44.2				Moderada	Pontadas
F	07/09/1966	G44.2	4	Anos	Holocrânia	Forte	Latejante
F	07/09/1966				Occipital/Nuca	Moderada	Peso ou Pressão
F	09/03/1982	G43.1			Holocrânia	Forte	Peso ou Pressão
F	09/03/1982					Moderada	
F	09/03/1982						
F	09/03/1982						
F	05/05/1973	F32.2	2	Anos	Hemocrânia	Forte	Peso ou Pressão
F	05/05/1973	G44.2			Frontal/Face	Moderada	
F	18/08/1981	G44.2			Holocrânia	Moderada	Peso ou Pressão
F	18/08/1981				Frontal/Face		

Surgimento	Evolução	Frequência	Duração	Sintomas Associados
		1 a 3 ao Mês	Horas	
Crises		1 a 3 ao Mês	Horas	Vômitos
Crises		Diária 1 a 2 por Semana +15 ao Mês	Horas Horas Horas	
		Diária		Tonturas
Contínua		Diária		Náuseas
		1 a 3 ao Mês	Horas	
Contínua	Estável	Diária	Horas	Náuseas
		Diária	Horas	
Crises		1 a 3 ao Mês	Horas	Fotofobia Fonofobia
Crises		1 a 3 ao Mês	Horas	Náuseas Fonofobia
EpIsódica/Periódica		1 a 2 por Semana	Horas	
EpIsódica/Periódica	Estável	1 a 2 por Semana	Horas	Náuseas Transtornos Visuais Tonturas
		+15 ao Mês	Horas	Náuseas
Contínua		Diária		
		1 a 2 por Semana	Horas	Náuseas
		+15 ao Mês	Horas Dias	Náuseas Fotofobia Fonofobia
		+15 ao Mês	Horas	Transtornos Visuais Náuseas
		+15 ao Mês		Náuseas

Agravantes	Atenuantes	Medicamento	Medicamentos
Nervosismo		Raramente	
Transtorno do Sono		Diário	Neosaldina
Nervosismo		Raramente	
Nervosismo		Raramente	
Nervosismo		Quase Diário	
Nervosismo			
Nervosismo		Sempre que Dói	Tylenol / Neosaldina
Nervosismo		Sempre que Dói Diário	
Transtorno do Sono		Sempre que Dói	Neosaldina / Aspirina
Esforço Físico			
Bebidas		Sempre que Dói	Neosaldina
Nervosismo	Deitar ou Dormir	Sempre que Dói	Tylenol
Bebidas		Sempre que Dói	Neosaldina
Nervosismo		Quase Diário	
Nervosismo		Raramente	
Menstruação			
Nervosismo		Sempre que Dói	
Nervosismo		Raramente	
Nervosismo		Raramente	
Nervosismo		Sempre que Dói	Tylenol
Nervosismo		Quase Diário	

...

ANEXO III

ENTRADAS DA RNA – CODIFICAÇÃO 1

2	3	4	3	3	1	2	1	1	3	1	1	2	9
1	3	4	2	8	1	0	0	1	0	0	2	0	9
1	4	4	2	8	2	0	0	1	0	6	1	0	9
1	3	4	3	8	1	2	1	1	3	6	25	0	11
1	3	4	1	1	1	1	1	4	4	49	0	0	11
1	4	4	6	2	8	1	1	1	3	1	1	6	6
1	3	4	9	3	2	0	0	1	0	49	10	0	6
1	3	0	6	8	4	0	0	2	3	40	9	0	6
1	3	3	2	5	5	0	0	1	0	49	9	0	6
1	3	4	2	8	8	2	1	1	3	11	9	0	6
1	3	4	2	8	5	0	0	2	3	1	9	0	6
1	3	4	2	8	7	0	0	1	0	1	9	0	6
1	4	4	15	3	2	2	0	1	3	0	2	0	6
1	4	4	3	8	2	2	0	1	0	0	2	0	6
1	3	3	2	3	2	2	0	1	0	49	1	0	6
1	3	3	2	5	2	3	0	2	0	6	1	0	6
1	3	4	3	8	1	0	0	1	0	6	1	0	6
2	4	4	2	0	3	0	0	2	3	5	1	0	6
1	3	3	3	1	2	2	0	1	0	15	1	0	6
1	3	4	6	8	2	0	0	1	4	1	1	0	6
1	3	4	2	8	1	0	0	2	3	1	1	0	6
1	3	4	3	8	1	0	0	2	0	1	1	0	6
1	3	2	2	1	7	0	0	1	0	1	1	0	6
1	3	4	15	8	2	0	0	1	0	1	1	0	6
1	4	4	2	8	2	0	0	1	0	1	1	0	6
1	3	4	2	8	4	3	0	5	5	0	1	0	6
1	4	4	2	8	2	2	0	1	5	0	1	0	6
2	4	4	3	8	2	3	0	3	4	0	1	0	6
2	4	0	1	3	3	0	0	2	4	0	1	0	6
2	4	4	2	3	2	0	0	2	3	0	1	0	6
1	4	4	2	8	2	2	0	1	3	0	1	0	6
1	4	4	2	2	1	0	0	1	3	0	1	0	6
1	3	3	9	1	1	2	0	1	0	0	1	0	6
1	4	4	1	8	14	0	0	1	0	0	1	0	6
1	3	4	2	1	2	3	0	2	3	6	0	0	6
2	4	3	2	8	2	0	0	1	3	6	0	0	6
1	3	4	2	8	2	2	0	1	0	2	0	0	6
1	4	4	1	1	1	1	0	2	4	1	0	0	6
1	4	3	6	1	7	2	0	1	0	1	0	0	6
1	3	4	0	1	5	1	0	2	9	0	0	0	6
2	3	4	20	1	8	1	0	2	3	0	0	0	6
1	4	0	1	8	2	2	0	1	3	0	0	0	6
2	4	3	1	8	2	0	0	1	3	0	0	0	6
1	4	4	1	8	4	2	0	1	0	0	0	0	6
2	4	3	2	3	2	2	0	1	0	0	0	0	6

...

0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	0	0	0	0	1	0	0	1	0	1	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
1	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	
0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	
1	0	0	0	0	1	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
1	0	0	0	1	0	0	0	0	1	1	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
1	0	0	0	1	0	0	0	0	0	1	0	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	1	0	1	0</															

[illegible]

...

ANEXO VI

EXEMPLO NUMÉRICO DE UMA REDE NEURAL

Dados os padrões do conjunto A com as seguintes entradas A1(0,2); A2(1,2); A3(1,3) que devem ter resposta igual a “1” e dados os padrões do conjunto B com as seguintes entradas B1(1,0); B2(2,1) que devem ter resposta igual a “0”, faça o treinamento de uma Rede Neural para executar a tarefa deste reconhecimento (padrão do conjunto A ou padrão do conjunto B) utilizando o algoritmo *backpropagation* em uma rede do tipo *feedforward* com duas camadas: a de entrada (deve ter 2 neurônios) e a camada de saída (com 1 neurônio). Faça os pesos iniciais serem todos iguais a zero e a taxa de aprendizagem $\gamma = 0,8$. Faça duas iterações completas, apresentando a análise da situação atual após cada iteração. Faça a interpretação geométrica.

Resolução:

Dados do problema:

$$w_1 = w_2 = \theta = 0$$

$$\gamma = 0,8$$

-----1ª ITERAÇÃO-----

Padrão A1 (0,2)

Propagação *forward*

$$i^{A1} = 0.0 + 0.2 + 0 = 0$$

$$a^{A1} = \frac{1}{1 + e^{-0}} = 0,5$$

Propagação *backward*

$$\delta^{A1} = 1 - 0,5 = 0,5$$

$$\Delta w_1 = 0,8.0.0,5 = 0$$

$$w_1(\text{novo}) = 0 + 0 = 0$$

$$\Delta w_2 = 0,8.2.0,5 = 0,8$$

$$w_2(\text{novo}) = 0 + 0,8 = 0,8$$

$$\Delta w_{\theta} = 0,8 \cdot 1,0,5 = 0,4$$

$$\theta(\text{novo}) = 0 + 0,4 = 0,4$$

Novos pesos:

$$w_1 = 0$$

$$w_2 = 0,8$$

$$\theta = 0,4$$

Padrão A2(1,2)

Propagação *forward*

$$i^{A2} = 0,1 + 0,8 \cdot 2 + 0,4 = 2,0$$

$$a^{A2} = \frac{1}{1 + e^{-2}} = 0,88$$

Propagação *backward*

$$\delta^{A2} = 1 - 0,88 = 0,12$$

$$\Delta w_1 = 0,8 \cdot 1,0,12 = 0,096$$

$$w_1(\text{novo}) = 0 + 0,096 = 0,096$$

$$\Delta w_2 = 0,8 \cdot 2,0,12 = 0,192$$

$$w_2(\text{novo}) = 0,8 + 0,192 = 0,992$$

$$\Delta w_{\theta} = 0,8 \cdot 1,0,12 = 0,096$$

$$\theta(\text{novo}) = 0,4 + 0,096 = 0,496$$

Novos pesos:

$$w_1 = 0,096$$

$$w_2 = 0,992$$

$$\theta = 0,496$$

Padrão A3(1,3)

Propagação *forward*

$$i^{A3} = 0,096 \cdot 1 + 0,992 \cdot 3 + 0,496 = 3,568$$

$$a^{A3} = \frac{1}{1 + e^{-3,568}} = 0,972$$

Propagação *backward*

$$\delta^{A3} = 1 - 0,972 = 0,0274$$

$$\Delta w_1 = 0,8 \cdot 1 \cdot 0,0274 = 0,022$$

$$w_1(\text{novo}) = 0,096 + 0,022 = 0,118$$

$$\Delta w_2 = 0,8 \cdot 3 \cdot 0,0274 = 0,066$$

$$w_2(\text{novo}) = 0,992 + 0,066 = 1,058$$

$$\Delta w_\theta = 0,8 \cdot 1 \cdot 0,0274 = 0,022$$

$$\theta(\text{novo}) = 0,496 + 0,022 = 0,518$$

Novos pesos:

$$w_1 = 0,118$$

$$w_2 = 1,058$$

$$\theta = 0,518$$

Padrão B1(1,0)

Propagação *forward*

$$i^{B1} = 0,118 \cdot 1 + 1,058 \cdot 0 + 0,518 = 0,636$$

$$a^{B1} = \frac{1}{1 + e^{-0,636}} = 0,654$$

Propagação *backward*

$$\delta^{B1} = 0 - 0,654 = -0,654$$

$$\Delta w_1 = 0,8 \cdot 1 \cdot (-0,654) = -0,523$$

$$w_1(\text{novo}) = 0,118 - 0,523 = -0,405$$

$$\Delta w_2 = 0,8 \cdot 0 \cdot (-0,654) = 0$$

$$w_2(\text{novo}) = 1,058 + 0 = 1,058$$

$$\Delta w_\theta = 0,8 \cdot 1 \cdot (-0,654) = -0,523$$

$$\theta(\text{novo}) = 0,518 - 0,523 = -0,005$$

Novos pesos:

$$w_1 = -0,405$$

$$w_2 = 1,058$$

$$\theta = -0,005$$

Padrão B2(2,1)**Propagação forward**

$$i^{B2} = -0,405 \cdot 2 + 1,058 \cdot 1 - 0,005 = 0,243$$

$$a^{B2} = \frac{1}{1 + e^{-0,243}} = 0,56$$

Propagação backward

$$\delta^{B2} = 0 - 0,56 = -0,56$$

$$\Delta w_1 = 0,8 \cdot 2 \cdot (-0,56) = -0,896$$

$$w_1(\text{novo}) = -0,405 - 0,896 = -1,301$$

$$\Delta w_2 = 0,8 \cdot 1 \cdot (-0,56) = -0,448$$

$$w_2(\text{novo}) = 1,058 - 0,448 = 0,61$$

$$\Delta w_\theta = 0,8 \cdot 1 \cdot (-0,56) = -0,448$$

$$\theta(\text{novo}) = -0,005 - 0,448 = -0,453$$

Novos pesos:

$$w_1 = -1,301$$

$$w_2 = 0,61$$

$$\theta = -0,453$$

----- FIM DA 1ª ITERAÇÃO -----

ANÁLISE DA SITUAÇÃO ATUAL:

Propagação forward utilizando os últimos pesos atualizados

Padrão A1 (0,2)

$$i^{A1} = -1,301 \cdot 0 + 0,61 \cdot 2 - 0,453 = 0,767$$

$$a^{A1} = \frac{1}{1 + e^{-0,767}} = 0,68$$

Padrão A2(1,2)

$$i^{A2} = -1,301.1 + 0,61.2 - 0,453 = -0,534$$

$$a^{A2} = \frac{1}{1 + e^{0,534}} = 0,37$$

Padrão A3(1,3)

$$i^{A3} = -1,301.1 + 0,61.3 - 0,453 = 0,076$$

$$a^{A3} = \frac{1}{1 + e^{-0,076}} = 0,52$$

Padrão B1(1,0)

$$i^{B1} = -1,301.1 + 0,61.0 - 0,453 = -1,754$$

$$a^{B1} = \frac{1}{1 + e^{1,754}} = 0,14$$

Padrão B2(2,1)

$$i^{B2} = -1,301.2 + 0,61.1 - 0,453 = -2,445$$

$$a^{B2} = \frac{1}{1 + e^{2,445}} = 0,079$$

Cálculo do Erro Quadrático:

$$E = \frac{1}{2} [(1 - 0,68)^2 + (1 - 0,37)^2 + (1 - 0,52)^2 + (0 - 0,14)^2 + (0 - 0,079)^2] = 0,337$$

-----2ª ITERAÇÃO-----

Novamente os padrões são apresentados à rede utilizando os últimos pesos atualizados para iniciar.

As saídas obtidas foram:

- $a^{A1} = 0,68$

- $a^{A2} = 0,67$
- $a^{A3} = 0,98$
- $a^{B1} = 0,28$
- $a^{B2} = 0,27$

Novos pesos:

$$w_1 = -1,69$$

$$w_2 = 1,46$$

$$\theta = -0,37$$

----- FIM DA 2ª ITERAÇÃO -----

ANÁLISE DA SITUAÇÃO ATUAL:

Propagação *forward* utilizando os últimos pesos atualizados

Padrão A1 (0,2)

$$a^{A1} = 0,93$$

Padrão A2(1,2)

$$a^{A2} = 0,70$$

Padrão A3(1,3)

$$a^{A3} = 0,91$$

Padrão B1(1,0)

$$a^{B1} = 0,11$$

Padrão B2(2,1)

$$a^{B1} = 0,09$$

Cálculo do Erro Quadrático:

$$E = 0,061$$

Interpretação Geométrica:

$$w_1x_1 + w_2x_2 + \theta = 0$$

$$-1,69x_1 + 1,46x_2 = 0,37$$

